ARL-TR-7849 ● OCT 2016

US Army Research Laboratory

Application of a Fuzzy Verification Technique for Assessment of the Weather Running Estimate–Nowcast (WRE–N) Model

by John W Raby

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-7849 ● OCT 2016

US Army Research Laboratory

Application of a Fuzzy Verification Technique for Assessment of the Weather Running Estimate–Nowcast (WRE–N) Model

by John W Raby
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
October 2016	Technical Report	October 2015–August 2016

4. TITLE AND SUBTITLE
Application of a Fuzzy Verification Technique for Assessment of the Weather Running Estimate–Nowcast (WRE–N) Model

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
John W Raby

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
Computational and Information Sciences Directorate
ATTN: RDRL-CIE-M
White Sands Missile Range, NM 88002

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7849

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Spatial forecasts from Numerical Weather Prediction (NWP) models of meteorological variables supporting US Army battlefield operations are an integral part of the products available for the Staff Weather Officer's use in providing mission forecasts. Army tactical decision aids (TDAs) ingest these forecasts; they are fused with information on weather thresholds, which impact the performance of systems and missions. This technical report presents some preliminary results obtained from the application of a nontraditional fuzzy verification method to evaluate the ability of NWP to simulate spatial variable fields filtered by the application of a threshold. Fuzzy methods have been developed in recent years to overcome limitations encountered when applying traditional verification techniques to high-resolution NWP forecasts, which often result in misleading assessments of forecast accuracy. This study illustrates how one fuzzy verification technique available from the Model Evaluation Tools Grid-Stat tool can be applied to the assessment of the Army Research Laboratory's Weather Running Estimate–Nowcast (WRE–N) model forecasts to which a threshold has been applied. Preliminary results suggest the fuzzy verification technique applied to the WRE–N provides an assessment of this unique aspect of model performance.

15. SUBJECT TERMS
fuzzy, neighborhood, weather impacts, TDA, thresholds, numerical weather prediction, observations, model verification, Model Evaluation Tools, assessment

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			John W Raby
Unclassified	Unclassified	Unclassified	UU	58	19b. TELEPHONE NUMBER (Include area code) 575-678-2004

Contents

List of Figures

List of Tables

Preface

This technical report relates to a previous work that explores the application of categorical and object-based verification methods to verify spatial forecasts produced by the Weather Running Estimate–Nowcast (WRE-N) of continuous meteorological variables that have been filtered by a threshold. These methods use gridded forecasts and observations on a common grid, which enables the application a number of different spatial verification methods that reveal various aspects of model performance. This report describes the results obtained when a different method called "fuzzy verification" was applied to the same data to determine its suitability in revealing more information about the ability of the WRE-N to predict objects defined by thresholds. Thus, portions of this report's content originated in ARL-TR-7751.[1]

[1] Raby JW, Cai H. Verification of spatial forecasts of continuous meteorological variables using categorical and object-based methods. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Aug. Report No.: ARL-TR-7751.

Acknowledgments

I offer my thanks to Mr Robert Dumais of the US Army Research Laboratory (ARL) who contributed guidance, data, suggestions, and information without which the study could not have been completed. I also thank Dr Huaqing Cai, Dr Brian Reen, and Dr Jeffrey Smith for their many suggestions for improvements in the application of the application of fuzzy verification techniques.

Many thanks go to Mr Martin Kufus of ARL Technical Publishing at White Sands Missile Range, New Mexico, for his consistently high standard of editing.

INTENTIONALLY LEFT BLANK.

Executive Summary

Spatial forecasts from Numerical Weather Prediction (NWP) models of tactically significant meteorological variables to support the US Army's battlefield operations have become an integral part of the products available for the Air Force Staff Weather Officer to use in providing mission planning and execution forecasts. Tactical decision aids (TDAs) ingest these forecasts. The TDAs fuse information on the characteristic operational weather thresholds that affect the performance of Army systems and missions with the NWP's spatial forecast information to generate spatial forecasts of these impacts for user-specified systems and/or missions for the time period and location of interest. This technical report presents methods that can be used to verify spatial-forecast fields of meteorological variables that have been filtered by the application of a threshold the same way as that used by the TDA. In effect, thresholds applied to a continuous variable field become categorical forecasts for which there are traditional and nontraditional methods for verification. This study evaluates the applicability of a nontraditional, fuzzy verification technique to assess unique aspects of model performance at domain-level in a way traditional techniques alone cannot.

Traditional grid-to-point methods can verify the skill of NWP in predicting continuous meteorological variables through the computation of such statistics as mean error and root-mean-square error, which characterize model accuracy over the entire domain. When these techniques are applied to high-resolution models such as the Army Weather Running Estimate–Nowcast (WRE–N), the results can give misleading error estimates when compared to lower-resolution models, which often score better when using these techniques. The issue is the inability of the verification technique to evaluate the true skill of higher-resolution forecasts, which replicate mesoscale atmospheric features in a way that is more representative of the actual phenomenon owing to their use of a reduced grid spacing over smaller domains, higher-resolution land-surface models, and better parameterization of subgrid physical processes.

In recent years, various nontraditional verification techniques have attempted to use different approaches to show the value of higher-resolution forecasts. In particular, spatial verification techniques have been developed that overcome the limitations of grid-to-point techniques, which score on the basis of the exact matching between point observations and the forecasts at those points. Fuzzy verification, also known as neighborhood verification, uses an approach that does not require exact matching but instead focuses on how well the atmospheric feature or object is replicated by the model, even if there is a spatial displacement of the feature. The goal is to

determine the amount of displacement by using a range of sizes of neighborhoods of surrounding forecast and observed grid points in the verification process. In this way, model performance as a function of spatial scale can be determined to allow selection of the scale required in order to have the desired accuracy. Many methods for fuzzy verification have been developed, mostly for evaluating model precipitation forecasts. Ebert reviews a number of such methods.[1] For this study, the minimum coverage method was applied to continuous meteorological variables.

The fuzzy verification framework used for this study requires the use of observations on a grid matching that of the WRE–N. Neighborhoods are defined in terms of the grid boxes within both grids. The number of grid boxes in one direction determines the size of the 2-D neighborhood or spatial scale when expressed in terms of the grid spacing for a particular model grid. For this study various spatial scales were chosen. The metrics and diagnostics used for scoring arise by defining an event from both the forecast and the observation grids. The event is defined by the use of a category or threshold as the basis for determining "hits" or "misses", which follows the established theoretical framework for evaluating deterministic binary forecasts. This framework evaluates the forecast skill by counting the numbers of times the event was forecast—or not—and observed—or not—in a contingency table. There are many statistics and skill scores that can be computed from the data collected by this method.

For this study, the authors obtained model output from the Army WRE–N, which is a version of the Advanced Research Weather Research and Forecasting Model adapted for generating short-range nowcasts and gridded observations produced by the National Oceanographic and Atmospheric Administration's Global Systems Division using the Local Analysis and Prediction System. A tool developed by the National Center for Atmospheric Research called MET Grid-Stat was used to apply the neighborhoods and thresholds to the grids and calculate the aggregate fuzzy-verification skill scores and contingency-table statistics for the entire domain.

Preliminary results suggest fuzzy verification scores and contingency-table statistics offer an assessment of a unique aspect of model performance which, when combined with the results of traditional methods and other nontraditional techniques, promises to provide a more comprehensive, domain-level assessment of model forecasts of continuous variables.

[1] Ebert E. Fuzzy verification of high resolution gridded forecasts: a review and proposed framework. Meteo App. 2008;15:51–64.

1. Introduction

As computing technology has advanced, the weather-forecasting task—once the job of a human forecaster in theater—has shifted to computerized Numerical Weather Prediction (NWP) models. Scientists around the world have used the Weather Research and Forecasting model (WRF) extensively for many applications. In this study, we have used the Advanced Research version of WRF (Skamarock et al. 2008) that we abbreviate as WRF–ARW. WRF–ARW includes Four-Dimensional Data Assimilation (FDDA) techniques that can incorporate observations into the model so that forecast quality is improved (Stauffer and Seaman 1994; Deng et al. 2009). The US Army Research Laboratory (ARL) uses WRF–ARW as the core of its Weather Running Estimate–Nowcast (WRE–N) weather forecasting model.

The Army requires high-resolution weather forecasting to model atmospheric features with wavelengths on the order of 5 km or less, which imposes a requirement for NWP to operate on a model grid spacing on the order of 1 km or less in the finest, or most resolved, domain in order to resolve weather phenomena of interest to the Soldier in theater. The atmospheric flows of interest to the Army include mountain/valley breezes, sea breezes, and other flows induced by differences in land-surface characteristics. High-resolution NWP forecasts need to be validated against observations before their outputs can be used by applications such as My Weather Impacts Decision Aid (MyWIDA), developed by Brandt et al. (2013). Weather forecast validation has always been of interest to the civilian and military weather forecasting community; see, for example, the reviews by Ebert et al. (2013) and Casati et al. (2008) or the guides by Jolliffe and Stephenson (2012) or Wilks (2011). The validation of the models, especially high-resolution NWP, has proven to be especially difficult when addressing small temporal and spatial scales (NRC 2010) that characterize NWP for use in Army applications. Furthermore, the verification of WRE–N spatial fields of continuous meteorological variables that have been filtered by the application of a threshold in order to evaluate the applicability of such output for use in MyWIDA has not been accomplished.

The WRF model is maintained by the National Center for Atmospheric Research (NCAR), which has also developed a suite of Model Evaluation Tools (MET) (NCAR 2013) to evaluate WRF–ARW performance. MET was developed at NCAR through a grant from the US Air Force 557th Weather Wing (formerly the Air Force Weather Agency). NCAR is sponsored by the US National Science Foundation. Grid-Stat, which is a tool in MET, provides verification statistics for matched forecast and observation grids. For fuzzy or neighborhood verification,

Grid-Stat compares the forecasts and observations at grid points in a neighborhood surrounding the point of interest rather than comparing a single point from both fields. For scoring, thresholds are set that establish the basis for defining forecast and observed events. Within a neighborhood, the number of observed events and forecast events are compared. By varying the neighborhood size, and rescoring, the relationship between neighborhood size and forecast skill can be determined. Grid-Stat performs this scoring at every grid point to calculate aggregate fuzzy-verification skill scores and contingency-table statistics for the entire domain.

ARL has employed the spatial verification tool MET Method for Object-Based Diagnostic Evaluation (MODE) in prior assessments such as that of Cai and Dumais (2015). They evaluated the 3-km grid spacing High Resolution Rapid Refresh model to demonstrate the utility of a nontraditional object-based technique in providing additional information to improve model precipitation forecasts to complement the information provided by the use of traditional verification techniques. In a separate study, Vaucher and Raby (2014) developed the capability to use MODE for object-based assessment of 1-km grid spacing WRE–N output of continuous meteorological variables. For this study, the only source of gridded observations available was from the National Oceanic and Atmospheric Administration (NOAA)–National Centers for Environmental Prediction Real-Time Mesoscale Analysis (RTMA) product (De Pondeca et al. 2011). In Vaucher and Raby (2014), the RTMA product, generated at a horizontal grid spacing of 2.5 km, was used with the WRE–N output that was remapped from a 1-km grid to a 2.5-km grid in order to produce the required matching grid.

MODE proved to be useful as an assessment tool for the WRE–N over an Army-scale domain and plans were made to expand its use to perform evaluations of continuous meteorological variables generated by the WRE–N at 1.75-km grid spacing. Collaborations with NOAA's Global Systems Division (GSD) resulted in the generation of 1.75-km grids of observations of surface meteorological variables for the same domain as the WRE–N using the NOAA–GSD Local Analysis and Prediction System (LAPS).

MET Series-Analysis was used in combination with MODE to perform spatial verification of the 1.75-km WRE–N by Raby and Cai (2016). This assessment demonstrated the value of combining the results from traditional categorical and nontraditional object-based verification methods for verification of the WRE–N. It also demonstrated how these methods verify spatial forecasts of continuous meteorological variables, which have been filtered by a single-threshold to quantify the degree of this particular type of accuracy, applicable to forecasts being used by the MyWIDA Tactical Decision Aid (TDA).

For this study, the WRE–N was run with FDDA for 5 case-study days over a 1.75-km grid-spacing domain in Southern California over highly varied terrain and with a dense observational network that provided a robust data set of model output for analysis. The case-study days from February through March 2012 were picked to vary weather conditions from a strong synoptic forcing situation to a quiescent situation. (The weather conditions for each study day are described in Subsection 2.3.)

This study illustrates how fuzzy verification techniques available from the MET Grid-Stat tool can be applied to the assessment of high-resolution WRE–N model forecasts. Fuzzy verification is a type of spatial verification that has been developed in recent years to address the inability of traditional verification techniques to adequately verify model forecasts, which are generated on increasingly smaller grids to resolve smaller-scale atmospheric features that are of interest to the Army. Traditional grid-to-point techniques score on the basis of the exact match between point observations and the forecasts at those points. When these techniques are applied to forecasts on grids with smaller spacing between grid points, the results are often misleading due to the error statistics being higher than those generated when the same technique is used for forecasts on grids whose points are spaced wider apart. In fact, the atmospheric features replicated by models using small-grid spacing bear more resemblance to the actual phenomena than those simulated at larger-grid spacings. The problem lies in requiring the exact match between the point observations and the forecast grid values. This leads to the so-called "double penalty" where the feature in the forecast being spatially displaced creates an offset in position that produces 2 types of errors. The first type results from the forecast placing the feature where it was not observed; the second type results from the forecast not placing the feature where it was observed. Furthermore, the error statistics provide no information about occurrences of "near-misses" that suggest a forecast of some quality or occurrences of more complete misses owing to a poor forecast. The challenge is to employ techniques that evaluate the ability of the model to replicate the features themselves, albeit with displacement, in addition to the more traditional objective approaches. To this end, researchers have developed spatial-verification techniques that reveal more about the ability of the model to predict spatial features (Jolliffe and Stephenson 2012).

Fuzzy verification uses an approach that does not require exact matching but instead focuses on how well the atmospheric feature or object is replicated by the model, even if there is a spatial displacement of the feature. The goal is to determine the amount of displacement by using a range of sizes of neighborhoods of surrounding forecast and observed grid points in the verification process. In this way, model performance as a function of spatial scale can be determined to allow

selection of the scale required in order to have the desired accuracy. Many methods for fuzzy verification have been developed, mostly for evaluating model-precipitation forecasts. Ebert (2008) reviews a number of such methods. For this study, the minimum coverage method was applied to continuous meteorological variables.

The minimum coverage method of fuzzy verification used for this study requires the use of observations on a grid matching that of the WRE–N. Neighborhoods are defined in terms of the grid boxes within both grids. The number of grid boxes in one direction determines the size of the 2-D neighborhood or spatial scale when expressed in terms of the grid spacing for a particular model grid. For this study various spatial scales were chosen. The metrics and diagnostics used for scoring arise by defining an event from both the forecast and the observation grids. The event is defined by the use of a category or threshold that serves as the basis for determining "hits" or "misses", which follows the established theoretical framework for evaluating deterministic binary forecasts. This framework evaluates the forecast skill by counting the numbers of times the event was forecast—or not—and observed—or not—in a contingency table. There are numerous statistics and skill scores that can be computed from the data collected by this method. When applied to a neighborhood, the fraction of grid squares within the neighborhood that contain events (i.e., the modeled and observed values met the threshold criterion) is compared with the total number of grid squares to derive a fractional coverage of events. This fractional coverage is then compared against the event-coverage threshold, which sets the minimum acceptable limit for deciding whether the model has scored a hit or not. This is done successively for each neighborhood size to generate the contingency-table statistics for each threshold and neighborhood size over the entire model domain.

In her review of fuzzy verification techniques Ebert (2008) points out the benefits of this approach in crediting forecasts that are close enough to show skill while providing additional information about the model, which can be used to improve the model. For example, fuzzy verification provides a measure of the quality of the forecast as a function of spatial scale through the use of neighborhoods. Another example is the ability to relate the skill of the forecast to the value of the threshold and the spatial scale in a way that allows one to identify the scale at which the model shows the desired level of skill. This information allows model developers to determine baseline levels of performance that can be compared with the same information generated following model upgrades to see details about how the upgrade improved the skill.

Jolliffe and Stephenson (2012) discuss the general expectations of model performance arising from the application of fuzzy verification methods to

precipitation forecasts. The first expectation is the forecast skill for low threshold values should exceed that for higher threshold values. The second is that the skill of the forecast should increase with increasing scale. Ebert (2008), using her case study for precipitation, relates the 2 by linking the highest skills with low threshold and large spatial scale. These expectations were applied in the analysis phase of this study, which involved continuous meteorological variables instead of precipitation to explore the applicability of fuzzy verification to continuous variable forecasts from the WRE–N.

2. Domain and Model

The ARL WRE–N (Dumais et al. 2004; Dumais et al. 2013) has been designed as a convection-allowing application of the WRF–ARW model (Skamarock et al. 2008) with an observation-nudging FDDA option (Liu et al. 2005; Deng et al. 2009). For this investigation, the WRE–N was configured to run over a multinest set of domains to produce a fine inner mesh with 1.75-km grid spacing and leveraged an external global model for cold-start initial conditions and time-dependent lateral boundary conditions for the outermost nest. Table 1 describes the dimensions for the triple-nested domain. This global model for ARL development and testing has been the National Center for Environmental Prediction's Global Forecast System (GFS) model (EMC 2003). The WRE–N is envisioned to be a rapid-update cycling application of WRF–ARW with FDDA and optimally could refresh itself at intervals up to hourly (dependent upon the observation network) (Dumais et al. 2012; Dumais and Reen 2013).

Table 1 WRE–N triple-nested domain dimensions in kilometers

East-West dimension	North-South dimension	Grid spacing
1780	1780	15.75
761	761	5.25
506	506	1.75

For this study, the model runs had a base time of 1200 coordinated universal time (UTC) and produced output for each hour from 1200 UTC to 0600 UTC of the following day for a total of 19 hourly model outputs, which were produced for each of 5 days in February and March 2012. The modeling domains are depicted in Fig. 1.

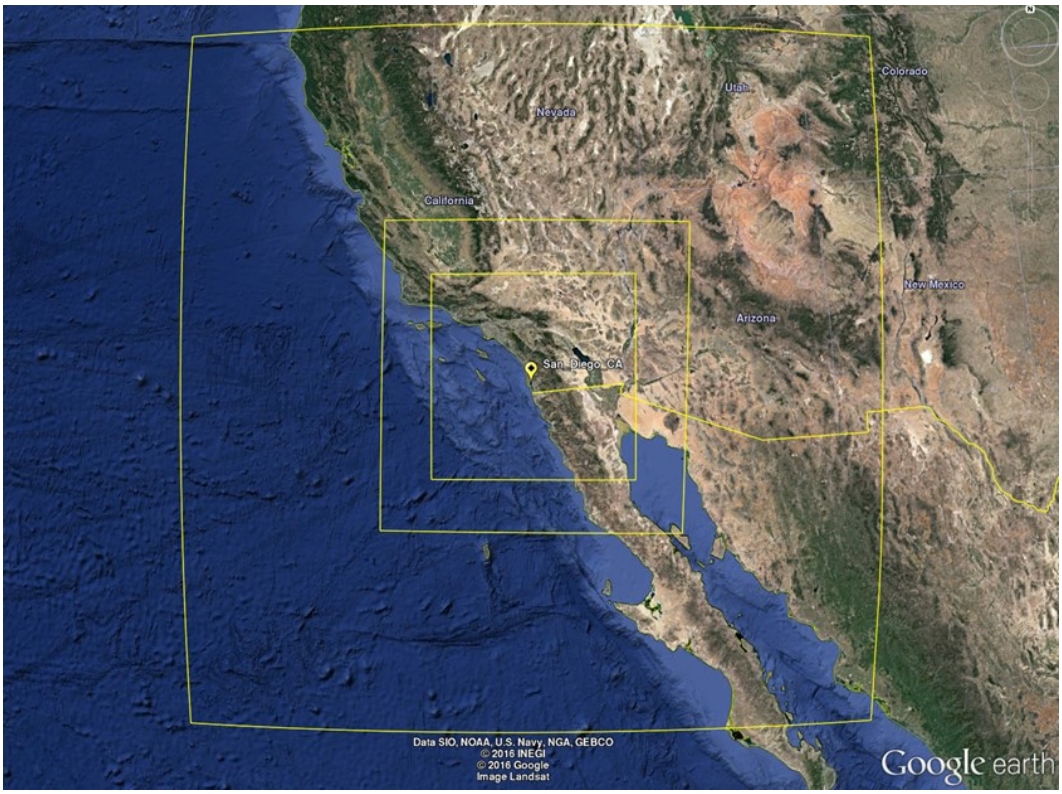

Fig. 1 Triple-nested model domains: domain center points are coincident and are centered near San Diego, California (Google Earth 2016)

2.1 Observations for Assimilation

The initial conditions were constructed by starting with the GFS data as the first guess for an analysis using observations. Most observations were obtained from the Meteorological Assimilation Data Ingest System (MADIS) (NOAA 2016), except for the Tropospheric Airborne Meteorological Data Reporting (TAMDAR) (Daniels et al. 2016) observations, which were obtained from AirDat, LLC. The MADIS database included standard surface observations, mesonet* surface observations, maritime surface observations, wind-profiler measurements, rawinsonde soundings and Aircraft Communications Addressing and Reporting System (ACARS) data. Use and reject lists were obtained from developers of the RTMA system (De Pondeca et al. 2011), and these were used to filter MADIS mesonet observations. This quality-assurance evaluation is especially important given the greater tendency of mesonet observations to be more poorly sited than other, more standard, surface observations.

The Obsgrid component of WRF was used for quality control of all observations. This included gross-error checks, comparison of observations to a background field

* A network of automated meteorological observation stations.

(here GFS), and comparison of observations to nearby observations. A modified version of Obsgrid allows for single-level observations such as the TAMDAR and ACARS data to be more effectively compared against the GFS background field. The quality-controlled observations were output in hourly, "little_r" formatted text files for use as ground-truth data for model assessment. Observation nudging was applied to the observations from these same sources for the preforecast period of 1200–1800 UTC (0- through 6-h lead times), followed by 1 h ramping down of the nudging from 1800 to 1900 UTC, during which no new observations are assimilated. The true, free forecast period thus begins at 1800 UTC because no observations after this time are assimilated.

2.2 Parameterizations

For the parameterization of turbulence in WRE–N, a modified version of the Mellor–Yamada–Janjić (MYJ) planetary boundary layer (PBL) (Janjić 1994) scheme was used. This modification decreases the background turbulent kinetic energy (TKE) and alters the diagnosis of the boundary-layer depth used for model output and data assimilation (Reen et al. 2014). The WRF single-moment, 5-class microphysics parameterization is used on all domains (Hong et al. 2004), while the Kain–Fritsch (Kain 2004) cumulus parameterization is used only on the 15.75-km outer domain. For radiation, the Rapid Radiative Transfer Model (RRTM) parameterization (Mlawer et al. 1997) is used for longwave radiation and the Dudhia (1989) scheme for shortwave radiation. The Noah land-surface model (Chen and Dudhia 2001a, 2001b) is used. Additional references and other details for these parameterization schemes are available from Skamarock et al. (2008). Table 2 lists the WRF configuration settings.

Table 2 WRE–N configuration

Configuration	Y/N?
WRF-ARW V3.4.1	Yes
Obs-nudging FDDA	Yes
Multinest (15.75/5.25/1.75km)	Yes
MADIS observations (FDDA)	Yes
TAMDAR observations (FDDA)	Yes
Ship/buoy observations (FDDA)	Yes
Filter obs (use/reject) (FDDA)	Yes
RUNWPSPLUS QC (FDDA)	Yes
Obs-nudge rad 120,60,20	Yes
MYJ-PBL scheme (modified)	Yes
WRF,sgl-moment, 5-class mp	Yes
Option 8 – microphysics	Yes
End FDDA 360 min	Yes
Kain-Fritsch Cum Param (outer dom)	Yes
RRTM longwave rad (Mlawer)	Yes
Shortwave rad (Dudhia)	Yes
Noah land surface model	Yes
Fix for nudge to low water vapor	Yes
Model Top 10 hPa	Yes
Feedback on	Yes
Obs weighting function 4E-4	Yes
57 vertical levels	Yes
48-s time step	Yes

2.3 Case-Study Days

The case-study days were selected on the basis of the prevailing synoptic weather conditions over the nested domains. Table 3 provides a short description of these conditions. For this study, results from analysis of Case 1 data only are presented.

Table 3 Synoptic conditions for the case-study days considered

Case	Dates (all 2012)	Description
1	February 07–08	Upper-level trough moved onshore, which led to widespread precipitation in the region.
2	February 09–10	Quiescent weather was in place with a 500-hPa ridge centered over central California at 1200 UTC.
3	February 16–17	An upper-level low located near the California–Arizona border with Mexico at 1200 UTC brought precipitation to that portion of the domain. This pattern moved south and east over the course of the day.
4	March 01–02	A weak shortwave trough resulted in precipitation in northern California at the beginning of the period that spread to Nevada, then moved southward and decreased in coverage.
5	March 05–06	Widespread high-level cloudiness due to weak upper-level low pressure but very limited precipitation.

2.4 Observations for Verification

The LAPS gridded observation data sets produced by NOAA–GSD consisted of 12 hourly Gridded Binary format, edition 2 (GRIB2) files of 2-m above-ground-level (AGL) temperature (TMP), relative humidity (RH), and dew-point temperature (DPT) and 10-m AGL U-component and V-component winds for the period of 1200–2300 UTC (forecast lead times 0 through 11) in each of the 5 cases. The output grid used by the LAPS was 289 × 289 with 1.75-km grid spacing. For this study, observations for 1200 UTC in the preforecast (assimilation) period and 1900 UTC in the forecast period (lead times 0 and 7, respectively) were used.

3. Data Preparation Using MET

The model and observational data were preprocessed into the formats required by MET Grid-Stat. The WRE–N model output data were converted from native Network Common Data Form files to hourly Gridded Binary format, edition 1 (GRIB) files by the WRF Unified Post Processor, which destaggers the data onto an Arakawa-A Grid containing 288 × 288 grid points. The hourly GRIB2 observation files on a 289 × 289 grid had to be remapped to the 288 × 288 grid to match that of the WRE–N grid. The NCAR "COPYGB" utility program was used to remap the observations and convert the files to GRIB format (DTC 2016). MET Grid-Stat was used to generate the grid-to-grid, fuzzy verification scores and contingency-table statistics aggregated over the entire 1.75-km WRE–N domain for surface meteorological variables TMP and DPT in degrees Kelvin (deg K), RH (%), and wind speed (WIND) in meters per second. Grid-Stat applied the specified neighborhood sizes (spatial scales) and thresholds and computed the neighborhood fractional coverage, contingency-table statistics, and skill scores for each spatial scale for 1200 UTC and 1900 UTC for Case 1. The event-coverage threshold used to decide whether the amount of fractional coverage for a neighborhood was considered a "hit" was greater than or equal to 0.5. The variable thresholds were specified using the FORTRAN convention of "GE" to indicate greater than or equal to the given threshold value and are shown in Table 4.

Table 4 Thresholds used in MET Grid-Stat

TMP (deg K)	DPT (deg K)	RH (%)	WIND (m/s)
265	262	25	2
270	267	40	5
275	272	55	8
280	277	70	11
285	282	85	14
290	…	…	…
295	…	…	…

The neighborhood sizes, in terms of number of grid squares (horizontal direction) and corresponding spatial scales for the 1.75-km grid spacing used in this study, are shown in Table 5.

Table 5 Neighborhood sizes and spatial scales (km) used in MET Grid-Stat

Grid squares	TMP scale	DPT scale	RH scale	WIND scale
1	1.75	1.75	1.75	1.75
3	5.25	5.25	5.25	5.25
5	8.75	8.75	8.75	8.75
7	12.25	12.25	12.25	12.25
9	15.75	15.75	15.75	15.75
11	19.25	…	…	…
13	22.75	…	…	…
15	26.25	…	…	…
17	29.75	…	…	…

MET Grid-Stat generates many fuzzy-verification skill scores and traditional contingency-table statistics. Of these, Table 6 lists those that were output initially for this study.

Table 6 Initial Grid-Stat fuzzy-verification skill scores and contingency-table statistics

Score/statistic	Description
FSS	Fractions skill score
BASER	Base rate
FMEAN	Mean forecast value
PODY	Hit rate
FAR	False-alarm ratio
FBIAS	Frequency bias
CSI	Critical success index
GSS	Gilbert Skill Score
ACC	Accuracy
MCTC	Multicategory Contingency Table Counts

The Grid-Stat output text files were ingested into Microsoft Excel spreadsheets, which were used to generate tabular and graphical displays showing how the scores and statistics, aggregated over the entire domain, vary with threshold value at different spatial scales. Text files containing the MCTC data were ingested into spreadsheets and plotted to show the distribution of counts of the forecast variable in various bins or ranges and the corresponding counts occurring in the same bins of the observed variable. This information is used to study model performance over discrete ranges of the variables. For this study, the analysis considered only CSI and FBIAS for the variables of 2-m AGL TMP and RH and 10-m AGL WIND—reducing the burden of analysis that comes when considering numerous scores and statistics during a preliminary evaluation of the applicability of the fuzzy-verification minimum coverage method in assessing the domain-level accuracy of WRE–N output filtered by the application of thresholds.

4. Analysis of MET Grid-State Fuzzy Verification Results

The CSI and FBIAS are defined by a ratio of counts determined using a 2 × 2 contingency table. Table 7 shows the contingency table with notation consistent with the formulae for the scores and statistics as implemented in the MET (NCAR 2013).

Table 7 The 2 × 2 contingency table from the MET User's Guide 4.1

2x2 contingency table in terms of counts. The n_{ij} values in the table represent the counts in each forecast-observation category, where i represents the forecast and j represents the observations. The "." symbols in the total cells represent sums across categories.

Forecast	Observation		Total
	o = 1 (e.g., "Yes")	*o* = 0 (e.g., "No")	
f = 1 (e.g., "Yes")	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
f = 0 (e.g., "No")	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
Total	$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	$T = n_{11} + n_{10} + n_{01} + n_{00}$

The counts, n_{11}, n_{10}, n_{01}, and n_{00}, are sometimes called the "Hits", "False alarms", "Misses", and "Correct rejections", respectively.

By dividing the counts in the cells by the overall total, T, the joint proportions, p_{11}, p_{10}, p_{01}, and p_{00} can be computed. Note that $p_{11} + p_{10} + p_{01} + p_{00} = 1$. Similarly, if the counts are divided by the row (column) totals, conditional proportions, based on the forecasts (observations) can be computed.

The CSI score (Eq. 1) is computed as described in the following excerpt from the MET User's Guide 4.1 (NCAR 2013):

$$\mathrm{CSI} = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}, \tag{1}$$

with CSI being the ratio of the number of times the event was correctly forecasted to occur to the number of times it was either forecasted or occurred. CSI ignores the "correct rejections" category (i.e., n_{00}).

The value of the CSI ranges between 0 and 1, with 1 being a perfect forecast and 0 being a forecast with no skill.

The FBIAS score is computed as described in Eq. 2:

$$\mathrm{Bias} = \frac{n_{11} + n_{10}}{n_{11} + n_{01}} = \frac{n_{1.}}{n_{.1}}, \tag{2}$$

with FBIAS defined as the ratio of the total number of forecasts of an event to the total number of observations of the event. A "good" value of frequency bias is close

to 1; a value greater than 1 indicates the event was forecasted too frequently and a value less than 1 indicates the event was not forecasted frequently enough.

4.1 Apply CSI for Fuzzy Verification of the WRE–N

The analysis of the CSI scores focuses on the forecast accuracy as defined by CSI as well as the degree to which the trend of skill, as a function of threshold value and spatial scale, follows the 2 expectations described in Section 1:

1) Skill increases with decreasing threshold value

2) Skill increases with increasing spatial scale

The ranges of the variables over the 1.75-km WRE–N domain at 1200 and 1900 UTC were used to establish the bounds within which the threshold values were selected. Initially, the range was divided into 4 intervals of equal size to determine the initial threshold values, with the highest value selected such that it would delineate an interval containing extreme values near the maximum value of the range. This simulated a situation similar to that which occurs when a MyWIDA TDA indicates an unfavorable impact based on forecast values exceeding the worst specific system or mission threshold.

Initial analyses of the CSI and FBIAS for TMP revealed a pattern of variability of these scores with threshold that varied little with spatial scale. It was decided the range of spatial scales should be expanded to include larger neighborhood sizes, capturing a stronger signal of the dependence of forecast skill on spatial scale. During a review of some early results that used the Fractions Skill Score, one reviewer suggested that expansion of the number of thresholds and the scales might reveal additional information about the relationship between the threshold and the scale (Smith 2016). Analysis of results for CSI and FBIAS produced with this expanded range of scales and larger number of thresholds failed to show a clear relationship between skill and scale. It was decided that for the other variables a reduced or nominal range of scales and number of thresholds would be sufficient to represent the full range of variation of scores with spatial scale for this study.

Scores were computed for a WRE–N valid time that fell in the preforecast period and a valid time that fell in the forecast period to provide a way to compare the forecast skill of the initialization grid of the GFS model interpolated onto the grid of the WRE–N with that of the WRE–N itself (Dumais 2016). Accordingly, the

model-output times selected for this comparison were the Hour 0 lead time for the GFS and the Hour 7 lead time for the WRE–N.*

Plots of the scores versus threshold at various spatial scales for the 3 variables TMP, RH, and WIND are now presented with an analysis of their characteristic features and their possible implications regarding model skill. In addition, 2-D maps of model TMP, RH, and WIND fields from the GFS (1200 UTC) and WRE–N (1900 UTC) were generated using a solid-color assignment to illustrate the spatial distribution and domain coverage of the variable where its value equals or exceeds each threshold. In further discussions regarding features in these maps, the term "object" will be used to refer to the areas shown where the value of the variable meets or exceeds the threshold. (Statements made regarding model performance only illustrate how an indication of performance can be gained from the scores and statistics taken alone. More comprehensive studies of large numbers of cases and additional scores and statistics are needed to verify model performance.)

Figure 2 is a display of CSI versus threshold value for the expanded range of spatial scales for TMP at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

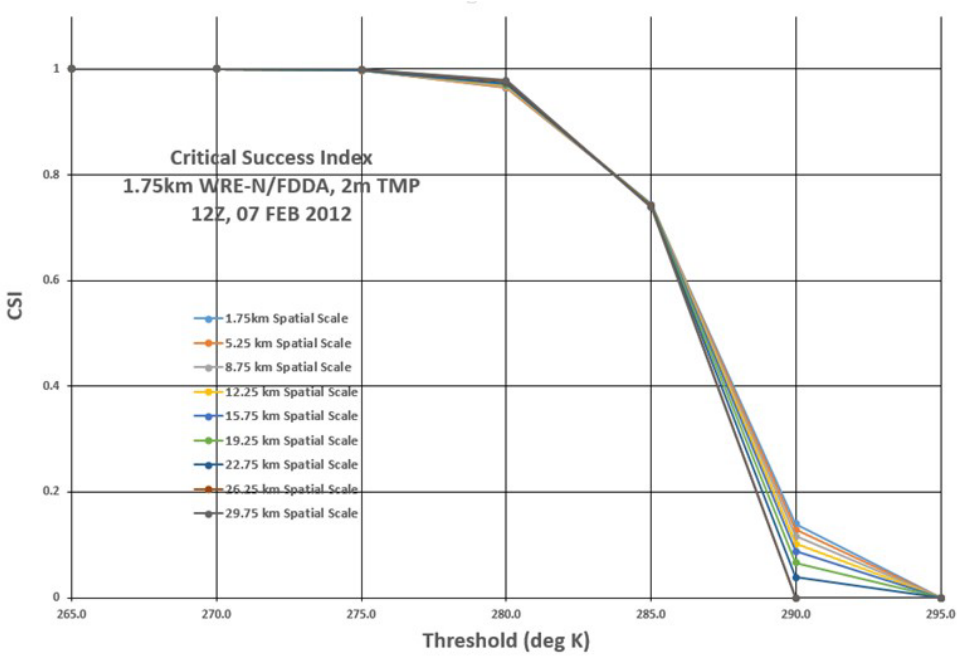

Fig. 2 CSI vs. threshold for a range of spatial scales for 2-m-AGL TMP at 1200 UTC for Case 1

At 1200 UTC the trend of decreasing CSI with increasing threshold value shows good skill for the GFS output mapped to the grid of the WRE–N for threshold

* Plots of the scores for Hour 0 lead time, although titled "WRE-N/FDDA", actually present the output of the GFS model interpolated onto the WRE–N grid.

values of 265–280 K, but the skill decreases markedly for higher threshold values, which is the expected trend overall. There is little evidence of the expected trend for which increases in the spatial scale result in an increase in the CSI, as noted by the curves coinciding for all threshold values—except for 290 K, where there is a very small spread in CSI values, with increasing CSI value with decreasing spatial scale, which contradicts the expected trend. It is likely this spread has little significance due to the relatively small range of CSI values compared to the range of scales.

Figure 3 is a map of the spatial distribution of GFS TMP color-shaded to depict the spatial distribution of the variable (object) defined where its value equals or exceeds each threshold at forecast valid time 1200 UTC for Case 1.

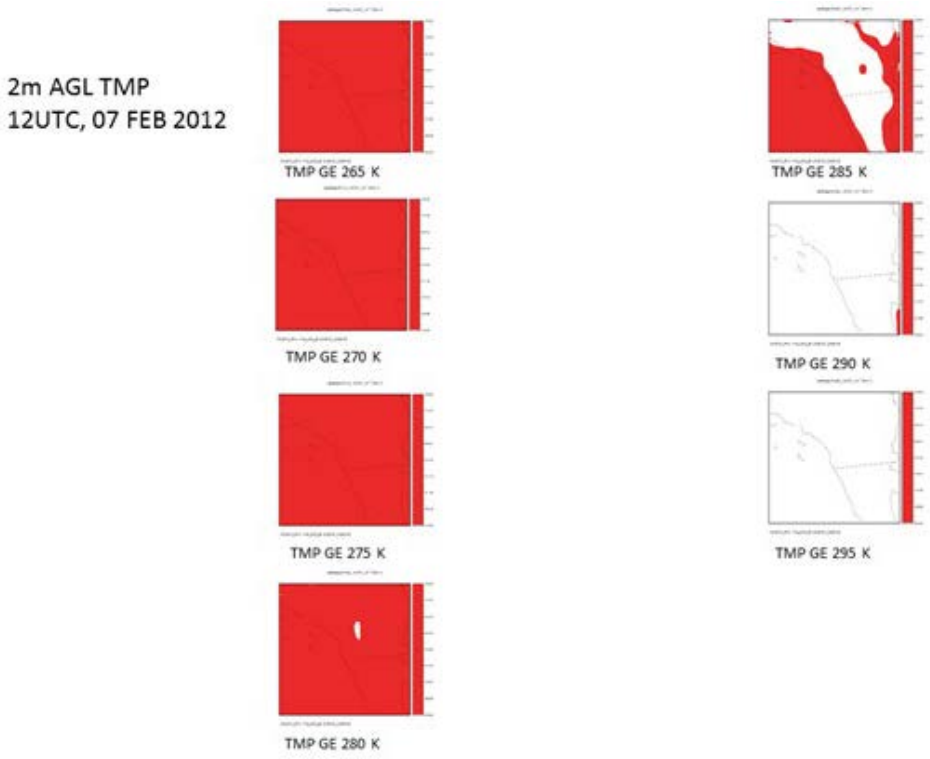

Fig. 3 Map of GFS 2-m-AGL TMP GE the 7 threshold values at 1200 UTC for Case 1

The spatial extent of the TMP object, as defined by the threshold for the first 3 thresholds, covers the entire domain. For thresholds of GE 280 K, the objects occupy a decreasing amount of the domain area. At GE 295 K, no object is present because no GFS TMP value equals or exceeds the threshold. The threshold value for which forecast skill, as indicated by CSI, begins a sharp decline (285 K) coincides with the threshold value for which the objects begin to decrease significantly in size. The possible relationship between object size and CSI for TMP may have implications for assessing the ability of the GFS to predict objects which,

in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting objects defined by the upper part of the range of the TMP.

Figure 4 displays CSI versus threshold value for the expanded range of spatial scales for WRE–N TMP at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

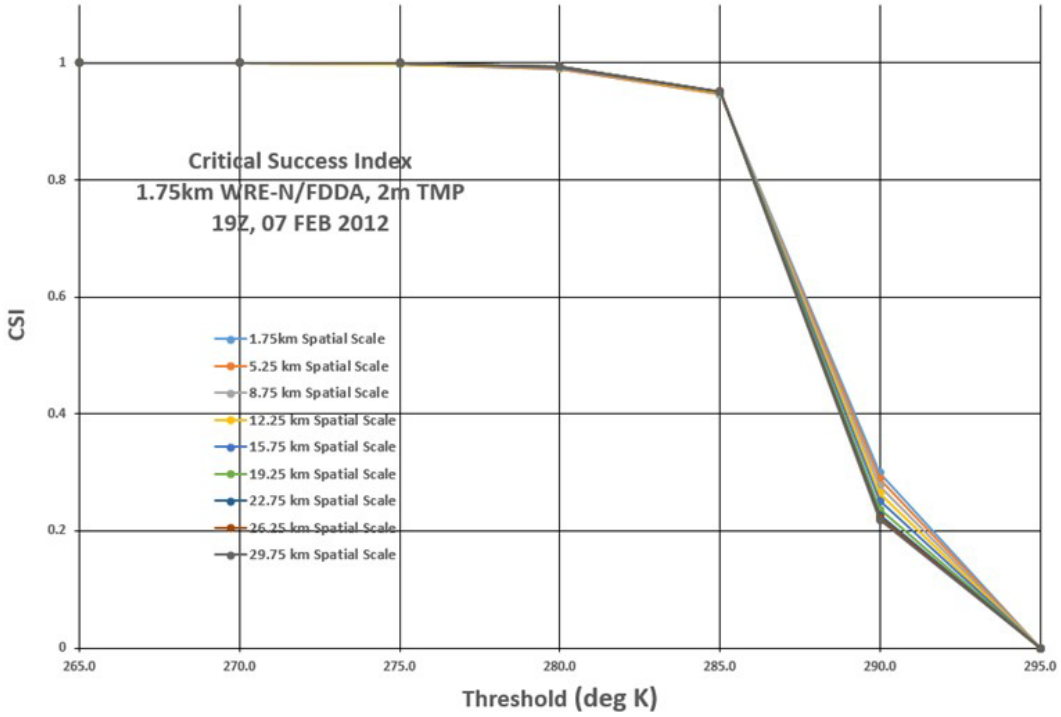

Fig. 4 CSI vs. threshold for a range of spatial scales for WRE–N 2-m-AGL TMP at 1900 UTC for Case 1

At 1900 UTC the same trend of CSI with threshold is present and agrees with the expected trend, but the sharp decrease of CSI with threshold starts at the higher threshold value of 290 K compared with 285 K at 1200 UTC. The range of threshold values over which the WRE–N shows good skill is larger than that of the GFS at 1200 UTC. There is no significant difference in CSI at a fixed threshold value within the range of spatial scales, which is counter to the expected trend.

Figure 5 maps the spatial distribution of WRE–N TMP color-shaded to depict the spatial distribution (object) of the variable defined where its value equals or exceeds each threshold at forecast valid time 1900 UTC for Case 1.

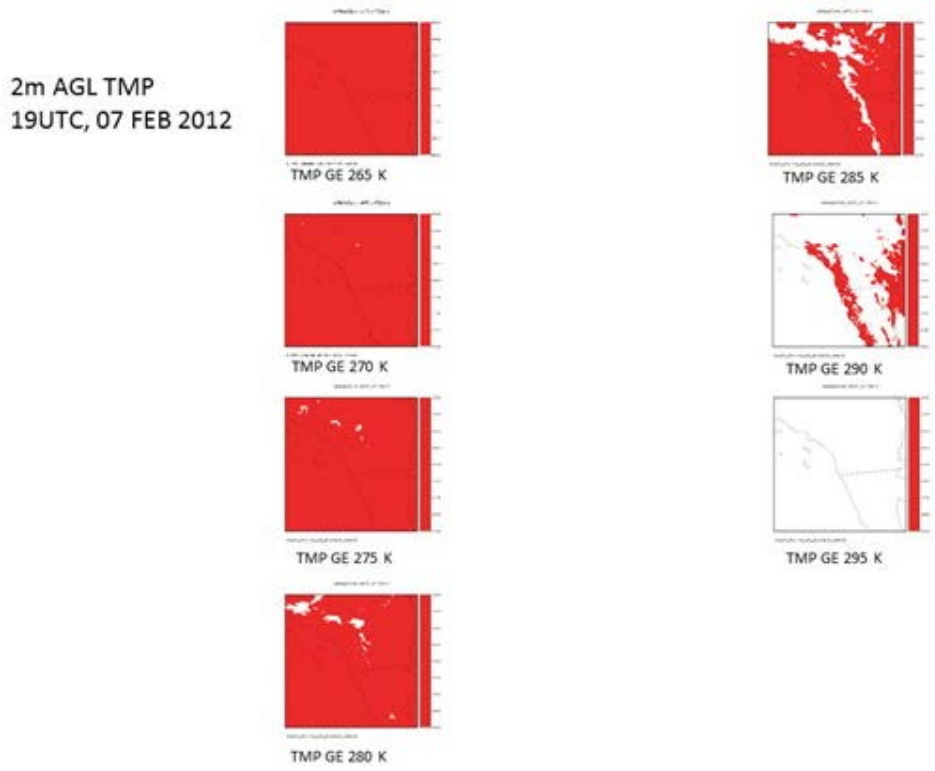

Fig. 5 Map of WRE–N 2-m-AGL TMP GE the 7 threshold values at 1900 UTC for Case 1

The spatial extent of the TMP object as defined by the first threshold (265 K) covers the entire domain. For thresholds of GE 270 K, the objects occupy a decreasing amount of the domain area. At GE 295 K, no object is present because no WRE–N TMP value equals or exceeds the threshold. The threshold value for which forecast skill, as indicated by CSI, begins a sharp decline (285 K) does not coincide with the threshold value of 270 K at which the objects begin to decrease in size. The object size at 285 K begins a more significant decrease with increasing threshold. The possible relationship between object size and CSI for TMP may have implications for assessing the ability of the WRE–N to predict objects which, in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting objects defined by the upper part of the range of the TMP.

Figure 6 displays CSI versus threshold value for 1.75-km spatial scale for TMP at valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

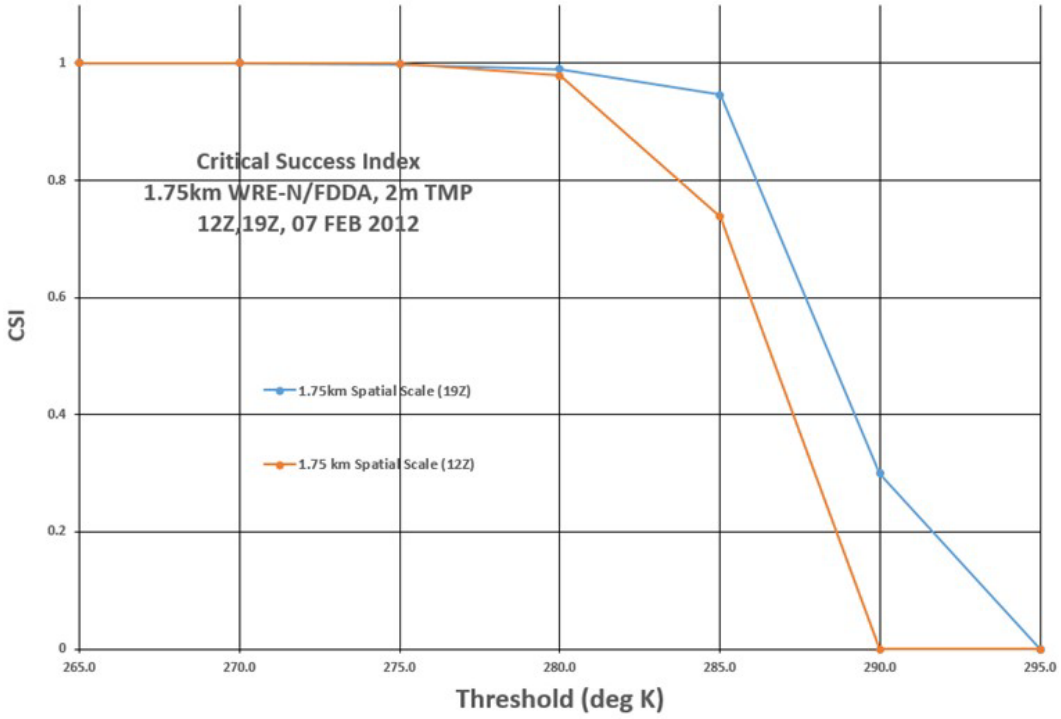

Fig. 6 CSI vs. threshold for 1.75-km spatial scale for 2-m-AGL TMP at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of CSI with threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The improved performance of the WRE–N over the GFS is more obvious with the break in high CSI occurring at a higher threshold for the WRE–N, which also shows more skill over a larger range of TMP. The WRE–N sustains its superior performance as the CSI scores drop with increasing threshold value.

Figure 7 displays CSI versus threshold value for the nominal range of spatial scales for GFS RH at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

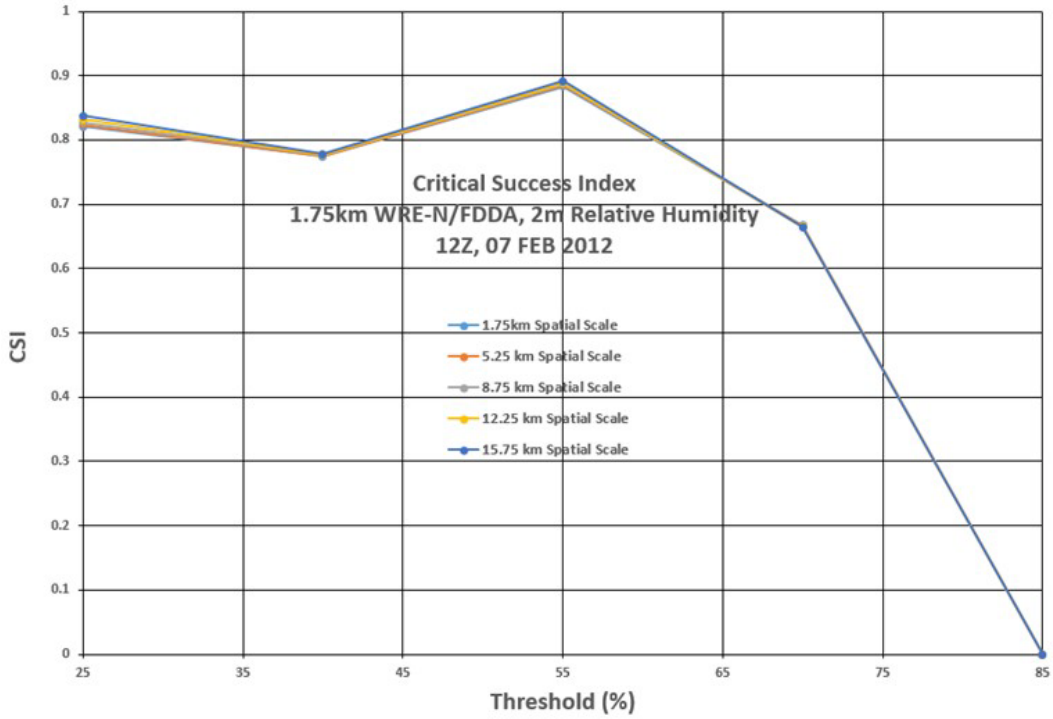

Fig. 7 CSI vs. threshold for a range of spatial scales for GFS 2-m-AGL RH at 1200 UTC for Case 1

The GFS model performs well at lower thresholds with a sharp decrease in CSI starting after 55%. Overall, the expected trend of CSI decreasing with increasing threshold value was confirmed. There is no significant difference in CSI at a fixed threshold value among the various spatial scales, which is not the expected trend. Over the range of thresholds where the CSI scores are good, the value of the score is not as high as that of TMP. This would indicate the GFS for TMP performs better than for RH.

Figure 8 maps the spatial distribution of GFS RH color-shaded to depict the spatial distribution (object) of the variable defined where its value equals or exceeds each threshold at forecast valid time 1200 UTC for Case 1.

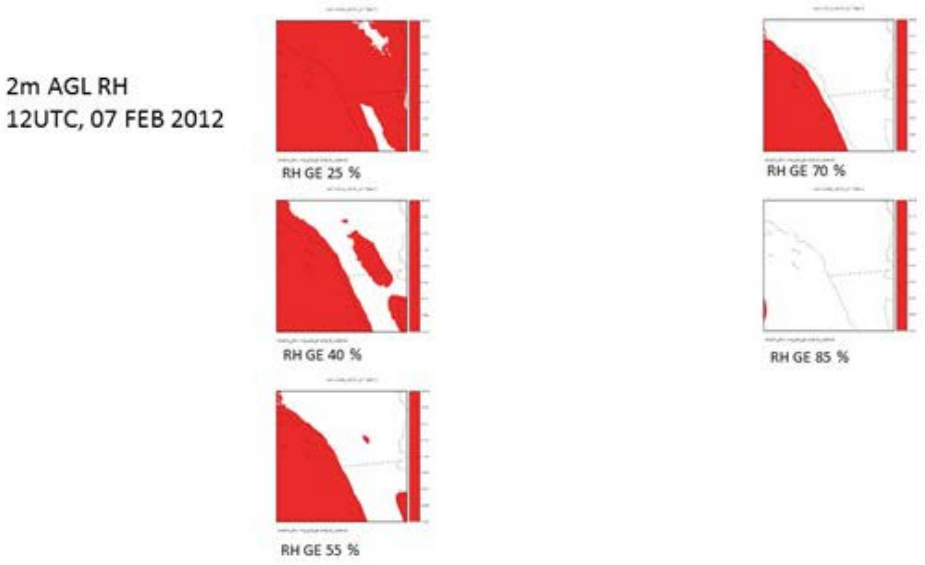

Fig. 8 Map of GFS 2-m-AGL RH GE the 5 threshold values at 1200 UTC for Case 1

The spatial extent of the RH object, as defined by the threshold for all 5 thresholds, covers a large portion of the domain at 25% and progressively decreases in size to almost no portion at 85%. The threshold value after which forecast skill, as indicated by CSI, begins a sharp decline (55%) coincides with the same threshold value for which the object is about half of the size of the domain. The possible relationship between object size and CSI for RH may have implications for assessing the ability of the WRE–N to predict objects which, in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting objects defined by the upper part of the range of the RH.

Figure 9 displays CSI versus threshold value for the nominal range of spatial scales for WRE–N RH at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

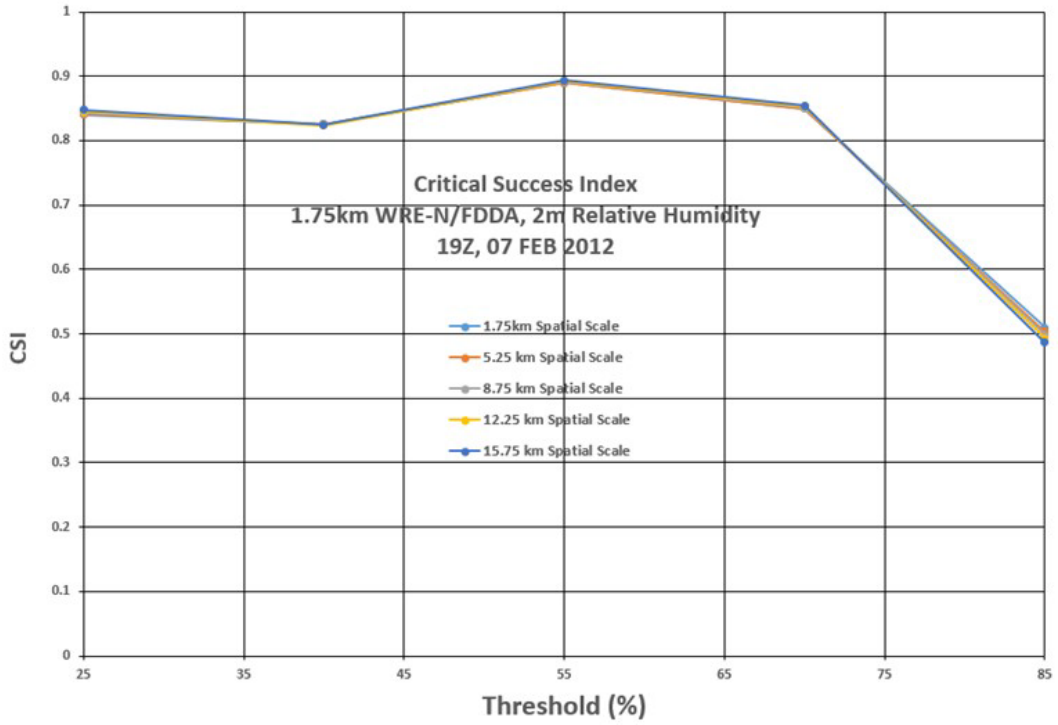

Fig. 9 CSI vs. threshold for a range of spatial scales for WRE–N 2-m-AGL RH at 1900 UTC for Case 1

At 1900 UTC, the WRE–N shows better scores over a wider range of thresholds than the GFS at 1200 UTC. The CSI scores are good between 25% and 70% then decrease with threshold after 70%, though never as low as that for TMP after its performance drops at higher threshold values. Overall, the plot shows the expected trend of CSI decreasing with increasing threshold value. There is no significant difference in CSI score at a fixed threshold value with spatial scale, which does not support the expected trend of increasing CSI with increasing spatial scale.

Figure 10 shows the spatial distribution of WRE–N RH color-shaded to depict the spatial distribution of the variable (object) defined where its value equals or exceeds each threshold at forecast valid time 1900 UTC for Case 1.

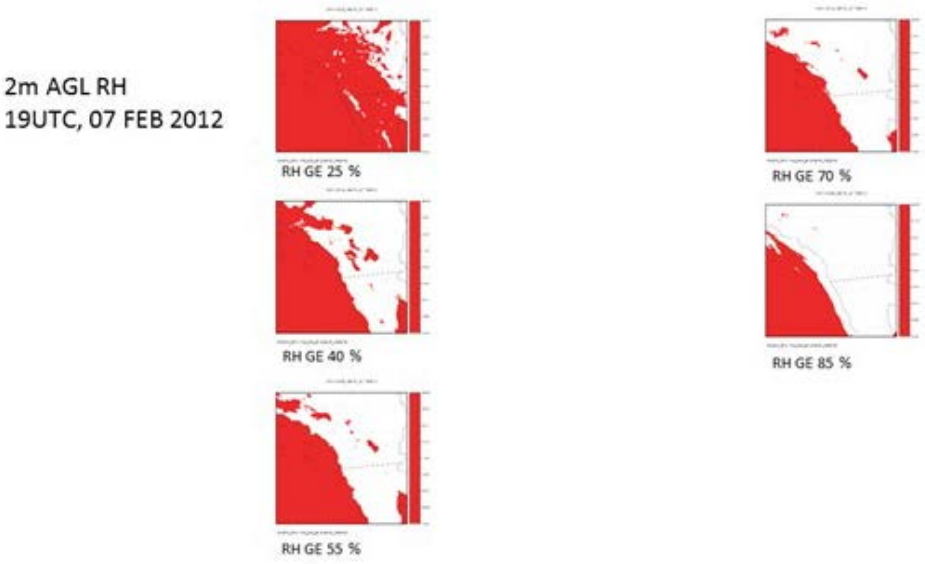

Fig. 10 Map of WRE–N 2-m-AGL RH GE the 5 threshold values at 1900 UTC for Case 1

The spatial extent of the RH objects for all 5 thresholds covers a large portion of the domain between 25% and 85%. Object size progressively decreases to less than half of the domain at 85%. The threshold value for which forecast skill, as indicated by CSI, begins a sharp decline (70%) coincides with the same threshold value for which the object is roughly half of the size of the domain. The possible relationship between object size and CSI for RH may have implications for assessing the ability of the WRE–N to predict objects which, in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting objects defined by the upper part of the range of the RH.

Figure 11 shows CSI versus threshold value for 1.75-km spatial scale for RH at valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

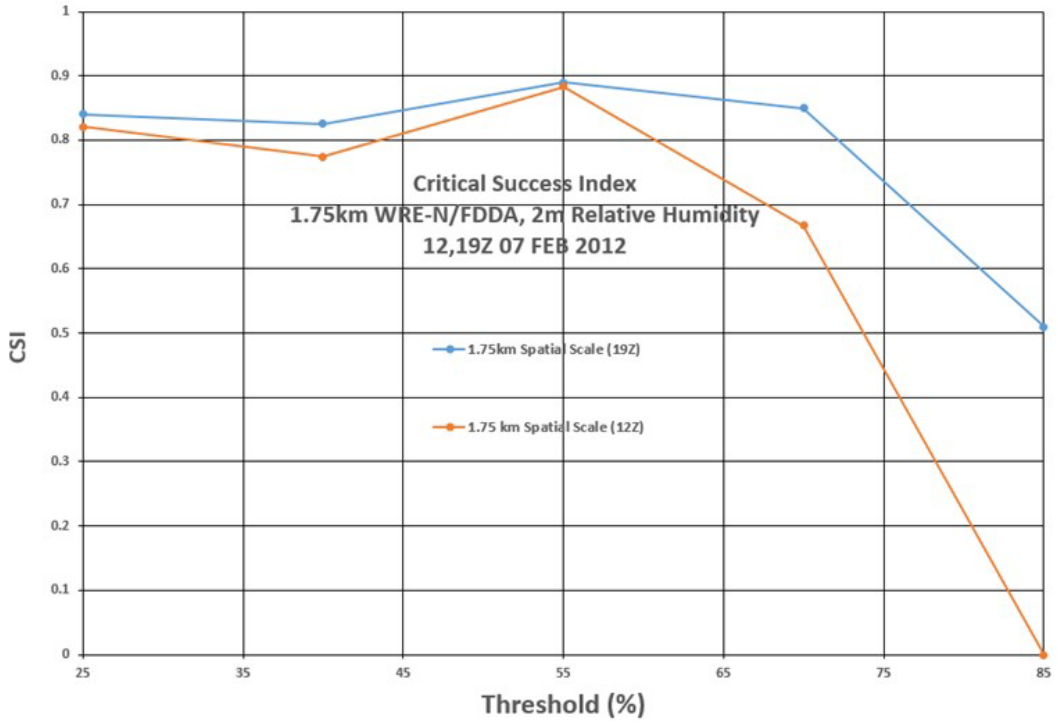

Fig. 11 CSI vs. threshold for 1.75-km spatial scale for 2-m-AGL RH at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of CSI with threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The CSI scores for the WRE–N at 1.75-km spatial scale are clearly superior to those of the GFS. The WRE–N sustains higher skill over a larger range of threshold values than that of the GFS. The reason for this is likely the larger object sizes at 1900 UTC, particularly at higher threshold values. At the highest threshold value, the CSI for the WRE–N never drops as low (0.5) as that of the GFS (0). For RH, the WRE–N shows reasonable skill at the highest threshold value, which was not the case for TMP where the skill was poor at the highest threshold; however, caution is advised on the significance of this because the size of the RH object at the highest threshold (roughly less than half of the domain) was significantly greater than that of the TMP object, which was 0 at the highest threshold.

Figure 12 displays CSI versus threshold value for the nominal range of spatial scales for GFS WIND at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

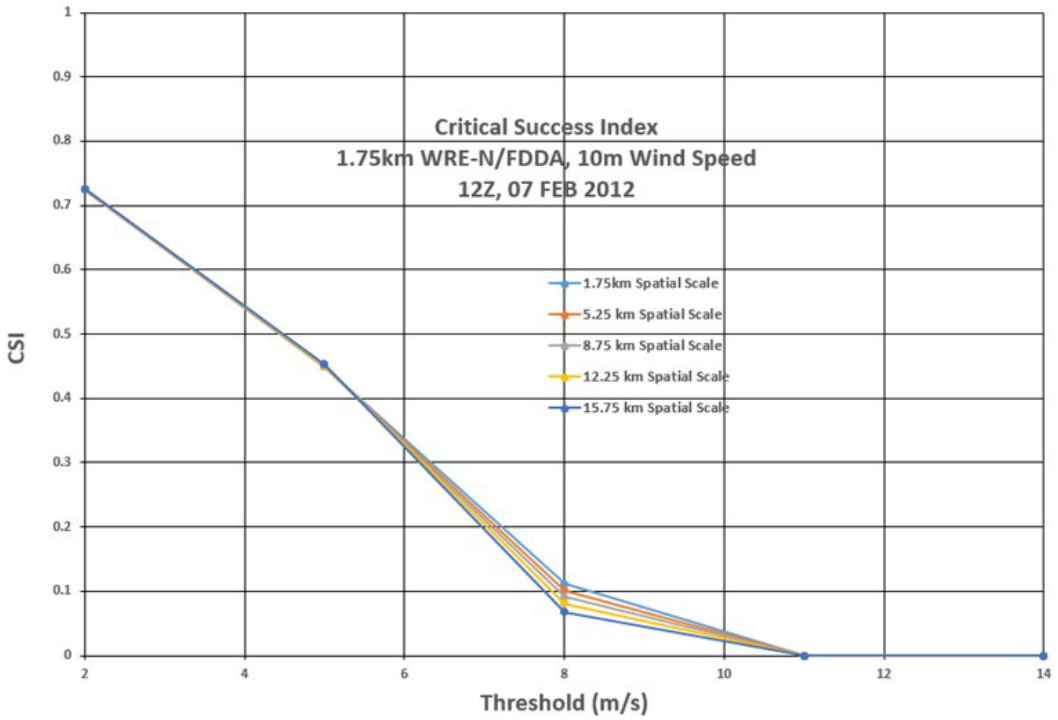

Fig. 12 CSI vs. threshold for a range of spatial scales for GFS 10-m-AGL WIND at 1200 UTC for Case 1

The GFS model, over the entire range of threshold values, performs less skillfully for WIND compared with the skill it had for TMP and RH. This is most notable at lower thresholds with a sharp decrease in CSI commencing immediately with increasing threshold value. The overall trend of CSI with threshold value is as expected with skill decreasing as the threshold increases. There is no significant difference in CSI at a fixed threshold value among the various spatial scales, which is not the expected trend. Overall, these results would indicate the GFS for TMP and RH performs better than for WIND.

Figure 13 shows the spatial distribution of model WIND color-shaded to depict the spatial distribution (object) of the variable defined where its value equals or exceeds each threshold at forecast valid time 1200 UTC for Case 1.

Fig. 13 Map of GFS 10-m-AGL WIND GE the 5 threshold values at 1200 UTC for Case 1

The spatial extent of the WIND object, as defined by the threshold for all 5 thresholds, covers a large portion of the domain at 2 m/s, rapidly decreases in size to a very small portion at 8 m/s, and further decreases to no object at higher threshold values. The threshold value for which forecast skill, as indicated by CSI, begins a sharp decline (2 m/s) coincides with the same threshold value for which the object occupies a sizeable portion of the domain. The possible relationship between object size and CSI for WIND may have implications for assessing the ability of the GFS to predict objects which, in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting WIND objects defined over the entire range of WIND.

Figure 14 shows CSI versus threshold value for the nominal range of spatial scales for WRE–N WIND at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

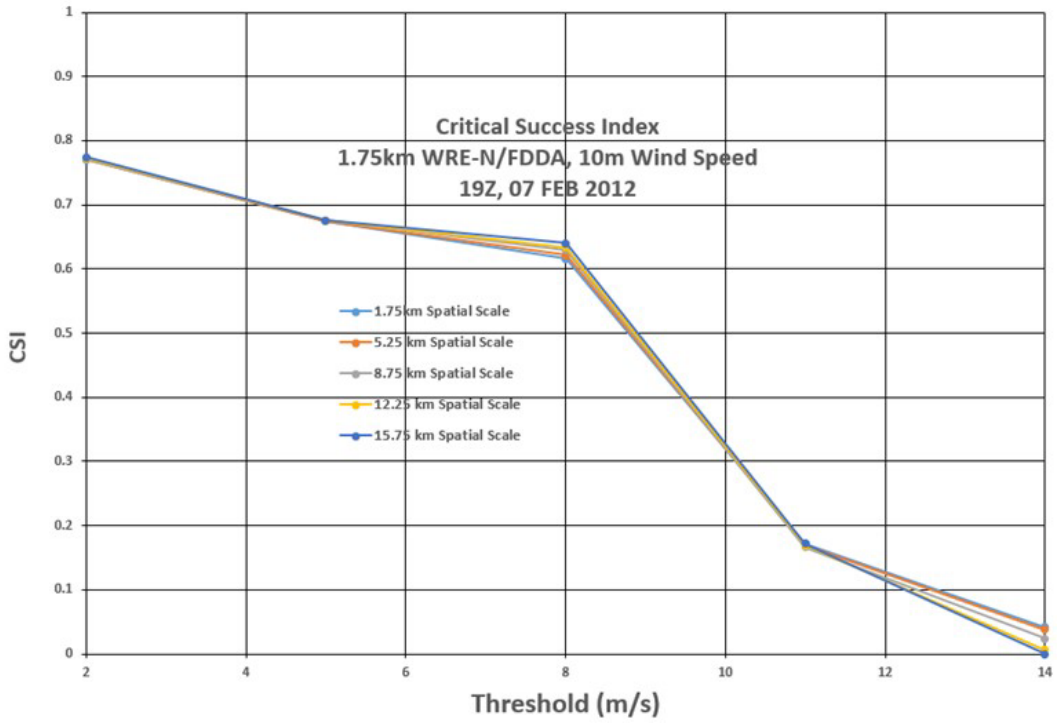

Fig. 14 CSI vs. threshold for a range of spatial scales for WRE–N 10-m-AGL WIND at 1900 UTC for Case 1

The WRE–N, over the entire range of threshold values, performs more skillfully for WIND compared with the GFS. The overall trend of decreasing CSI with increasing threshold value is as expected. However, the overall skill for WIND is less than that for TMP and RH. Unlike the GFS, the CSI does not drop significantly until after the threshold reaches a value of 8 m/s. There is no significant difference in CSI at a fixed threshold value among the various spatial scales, which is not as expected. These results would indicate the WRE–N for TMP and RH performs better than for WIND.

Figure 15 depicts the spatial distribution of model WIND color-shaded to show the spatial distribution of the variable (object) defined where its value equals or exceeds each threshold at forecast valid time 1900 UTC for Case 1.

Fig. 15 Map of WRE–N 10-m-AGL WIND GE the 5 threshold values at 1900 UTC for Case 1

The spatial extent of the WIND object, as defined by the threshold for all 5 thresholds, covers a large portion of the domain at 2 m/s and decreases in size to a very small portion at the highest threshold value of 14 m/s. The threshold value after which forecast skill, as indicated by CSI, begins a sharp decline (8 m/s) that coincides with the same threshold value for which the object occupies an area roughly half of the size of the domain. The possible relationship between object size and CSI for WIND may have implications for assessing the ability of the WRE–N to predict objects which, in turn, impacts the input data used by MyWIDA. One factor that may contribute to this decrease in skill is the smaller objects, which are defined at higher threshold values. Matching for small objects between the forecast and the observations tends to be difficult because it requires a smaller displacement error. Analysis of more data is needed to confirm this apparent loss of skill when forecasting WIND objects defined over the higher portion of the range of WIND.

Figure 16 displays CSI versus threshold value for 1.75-km spatial scale for WIND at valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

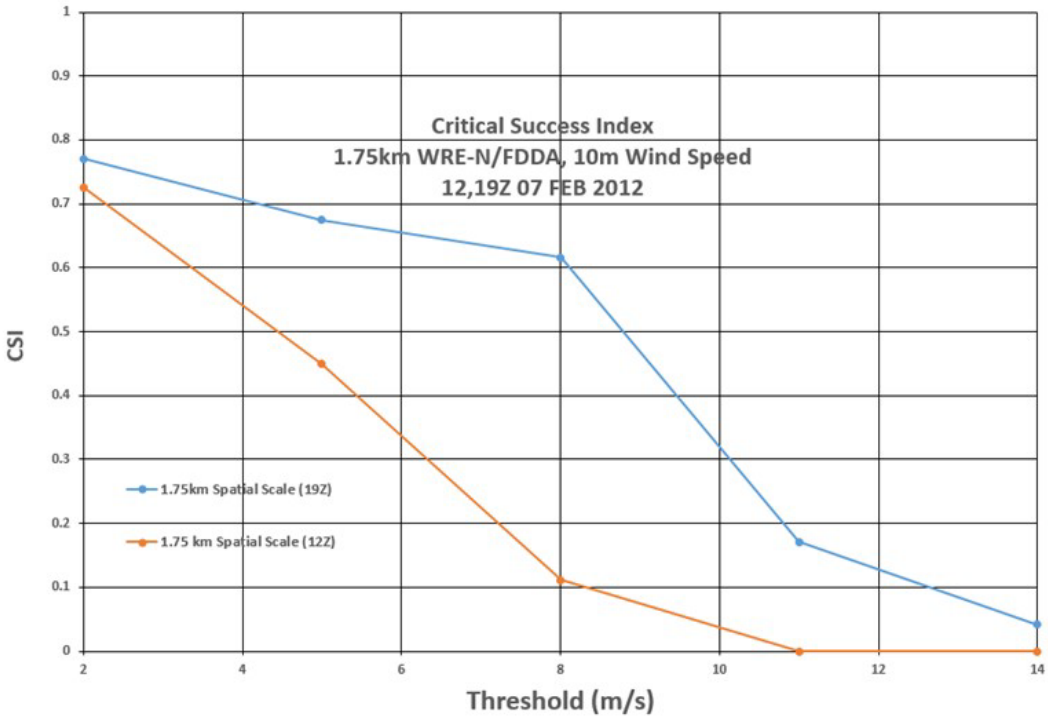

Fig. 16 CSI vs. threshold for 1.75-km spatial scale for 10-m-AGL WIND at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of CSI with the threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The CSI scores for the WRE–N at 1.75-km spatial scale are clearly superior to those of the GFS. The WRE–N sustains higher skill over a larger range of threshold values than that of the GFS. Between the threshold of 2 and 8 m/s, the superior skill of the WRE–N is clearly shown. The reason for this is likely the larger object sizes at 1900 UTC, particularly between 2 and 8 m/s when compared with the sizes at 1200 UTC. Both models show poorer performance above 8 m/s as object sizes are reduced considerably.

4.2 Apply FBIAS for Fuzzy Verification of the WRE–N

The analysis of the FBIAS scores will focus on the forecast accuracy, as defined by FBIAS, as well as the degree to which the trend of skill as a function of threshold value and spatial scale follows the expectations described in Section 1:

1) Bias increases with increasing threshold value

2) Bias decreases with increasing spatial scale

Figure 17 shows FBIAS versus threshold value for the expanded range of spatial scales for GFS TMP at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

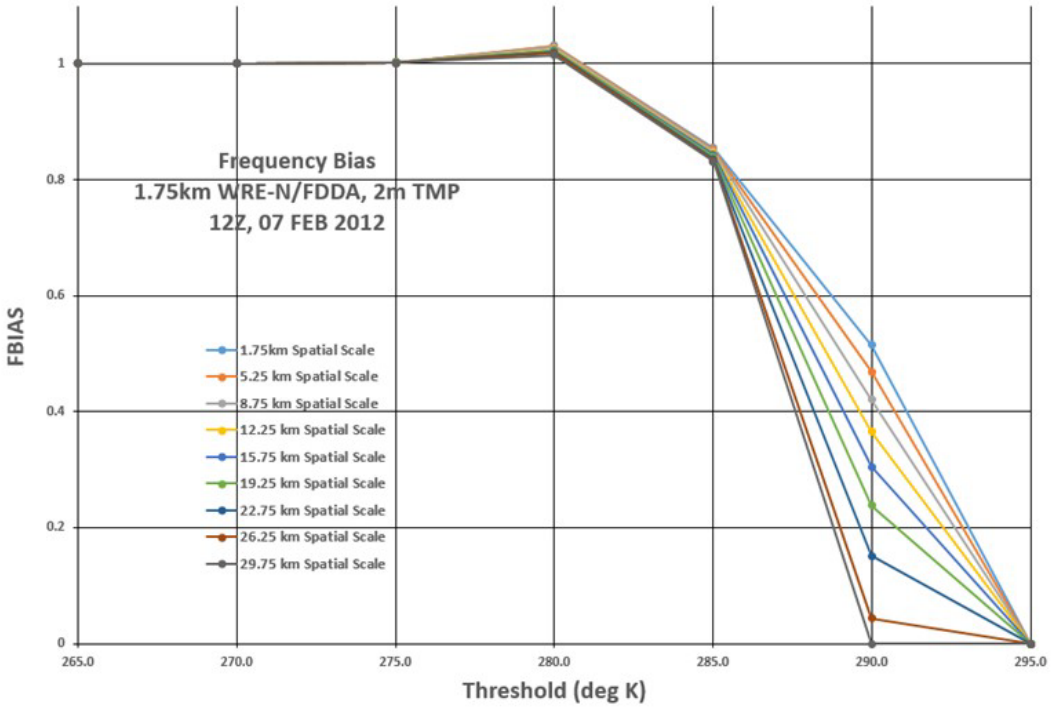

Fig. 17 FBIAS vs. threshold for a range of spatial scales for GFS 2-m-AGL TMP at 1200 UTC for Case 1

The FBIAS trend with increasing threshold value is as expected and is similar to that of CSI for TMP with regard to the threshold value of 280 K being the breakpoint between good and poorer performance. For FBIAS, the GFS shows almost no bias between 265 and 275 K; then, there is a small over-forecast bias at 280 K becoming an increasingly significant under-forecast bias at higher thresholds. There is little evidence of the expected trend—increases in the spatial scale result in a decreases in the FBIAS—as noted by the curves coinciding for all threshold values. The exception is 290 K, where there is some evidence of a dependence of bias value on spatial scale, which shows the 1.75-km scale with a lower bias than the 29.75-km scale, but the relationship is reversed from that expected. The significance of this and the amount of spread is worth further investigation since the range of bias values compared to the range of scales is not as small as was found for CSI at the same threshold. The improved bias with smaller scale does suggest better performance at smaller grid spacings; but why this dependency on scale occurs only at 290 K and why it does not follow the expected trend of skill increasing with increasing scale needs further explanation. From

Fig. 3, the 280-K breakpoint in FBIAS coincides with the threshold where object size commences a significant decrease with increasing threshold value.

Figure 18 displays FBIAS versus threshold value for the expanded range of spatial scales for WRE–N TMP at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

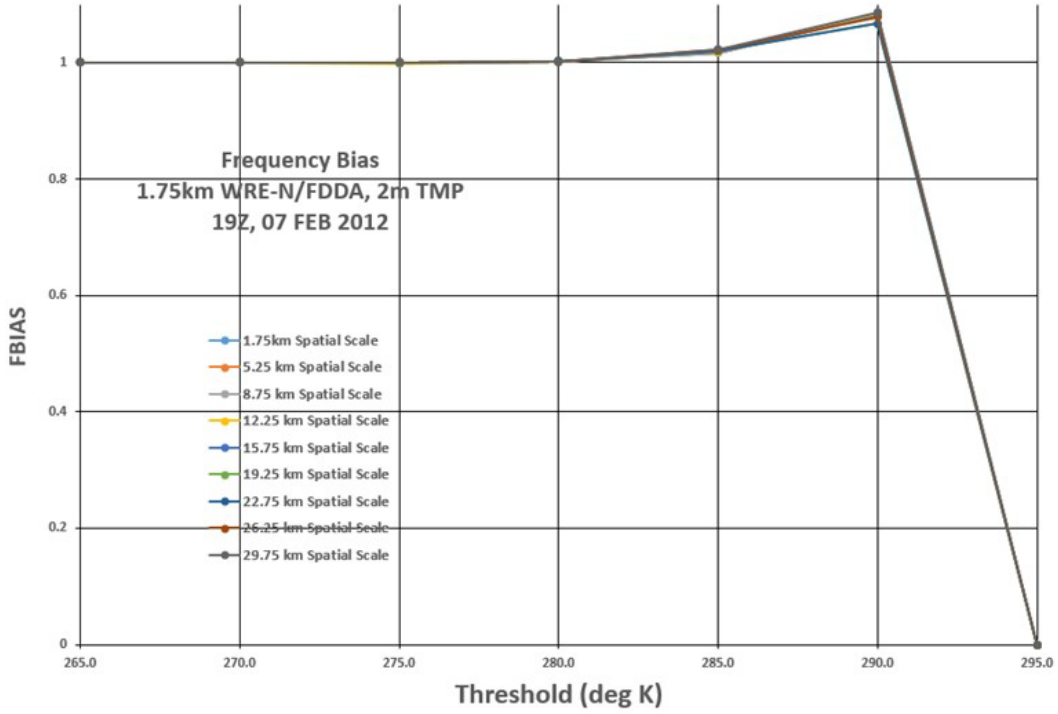

Fig. 18 FBIAS vs. threshold for a range of spatial scales for WRE–N 2-m-AGL TMP at 1900 UTC for Case 1

For the WRE-N, the FBIAS trend with increasing threshold value is expected and similar to that of CSI for TMP, but the breakpoint-threshold value of 280 K (between good performance and poorer performance) does not coincide as well as that for the GFS. For FBIAS, the WRE–N shows almost no bias between 265 and 285 K; then, a small over-forecast bias is evident at 290 K, switching to a significant under-forecast bias at the highest threshold of 295 K. There is little evidence of the expected trend—increases in the spatial scale result in a decrease in the FBIAS—as noted by the curves coinciding for all threshold values. From Fig. 5, the validity of such a large under-forecast bias at 295 K is questionable since there is no object defined at that threshold.

Figure 19 shows FBIAS versus threshold value for 1.75-km spatial scale for TMP at model valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

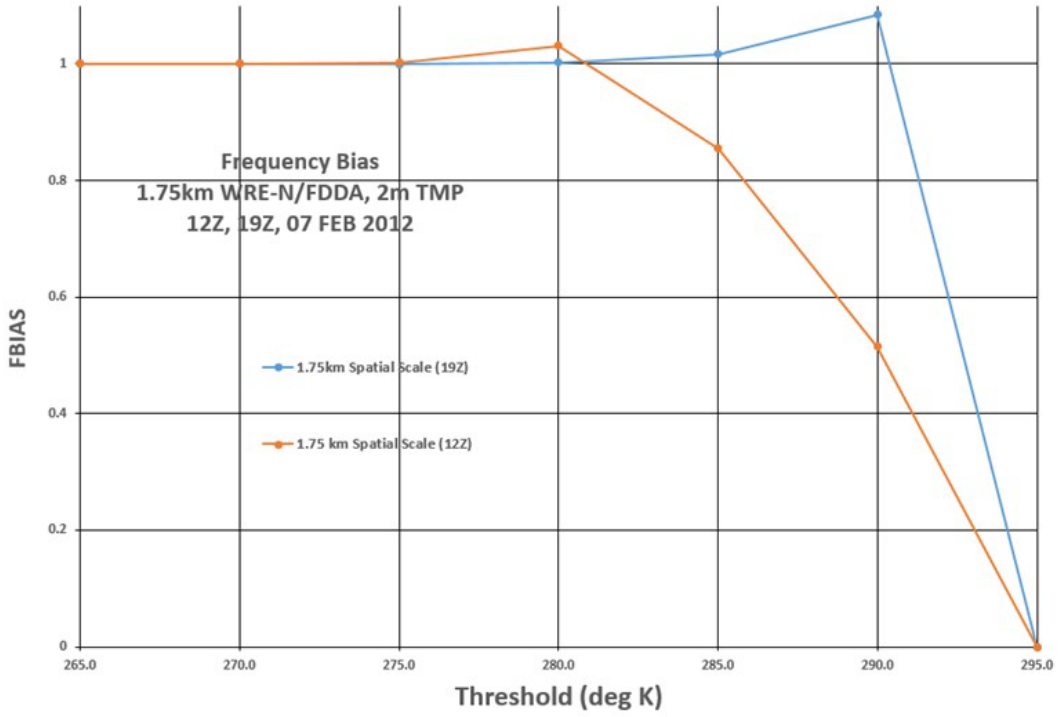

Fig. 19 FBIAS vs. threshold for 1.75-km spatial scale for 2-m-AGL TMP at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of FBIAS for TMP with threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The improved performance of the WRE–N over the GFS is more obvious with the break toward slight over-forecast occurring at a higher threshold for the WRE–N, which also shows less bias over a larger range of TMP.

Figure 20 displays FBIAS versus threshold value for the nominal range of spatial scales for GFS RH at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

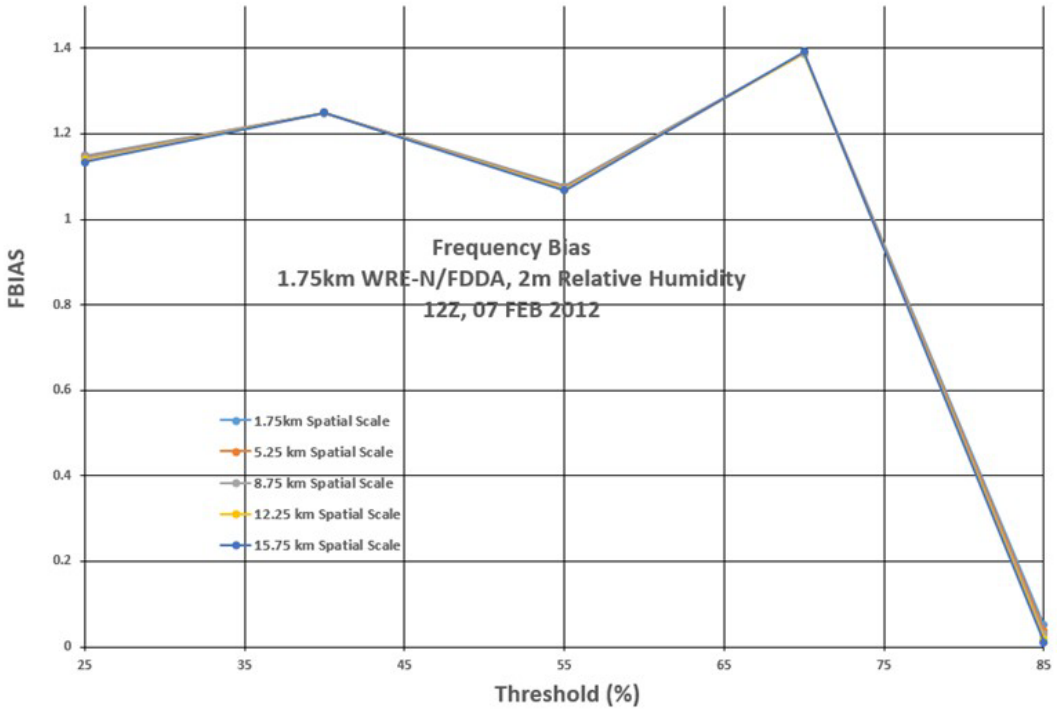

Fig. 20 FBIAS vs. threshold for a range of spatial scales for GFS 2-m-AGL RH at 1200 UTC for Case 1

The RH FBIAS for the GFS shows a consistent, moderate over-forecasting trend for all threshold values from 25% to 70% followed by a reversal to under-forecast at the highest threshold of 85%. This reversal and the extremely negative bias value of 0 at 85% cast doubt on the validity of this FBIAS value. The expected trend of decreasing bias with decreasing threshold value is not strictly present between 25% and 55%. From Fig. 8, note the size of the object at the same threshold (85%) is extremely small, which may be influencing the calculation of FBIAS. There is little evidence of the expected trend—increases in the spatial scale result in a decrease in the FBIAS—as noted by the curves coinciding for all threshold values. More investigation is needed on the impact of very small objects on the validity of FBIAS values.

Figure 21 shows FBIAS versus threshold value for the nominal range of spatial scales for WRE–N RH at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

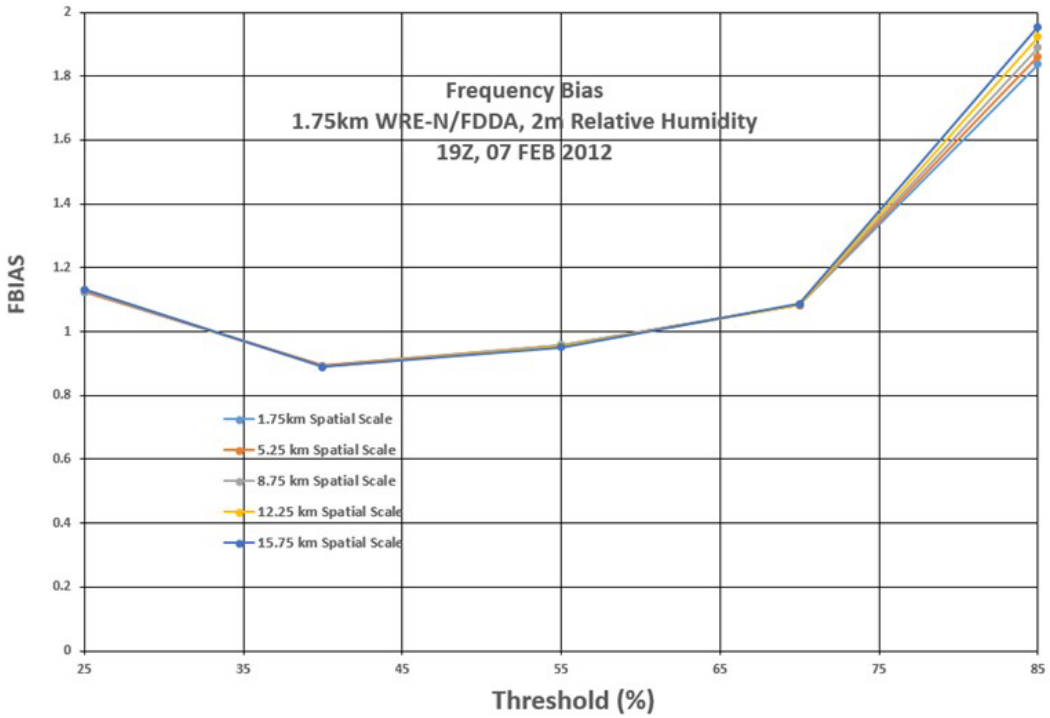

Fig. 21 FBIAS vs. threshold for a range of spatial scales for WRE–N 2-m-AGL RH at 1900 UTC for Case 1

The RH FBIAS for the WRE–N shows little or no bias for all threshold value from 25% to 70% followed by a decided trend toward over-forecast at the highest threshold of 85%. This trend of bias is as expected. The large positive bias value of about 1.9 at 85% is noteworthy. From Fig. 10, note that the size of the object at the same threshold is almost half of the domain, which does not suggest that object size is an issue. There is little evidence of the expected trend—increases in the spatial scale result in a decrease in the FBIAS—as noted by the curves coinciding for all threshold values.

Figure 22 displays FBIAS versus threshold value for 1.75-km spatial scale for RH at valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

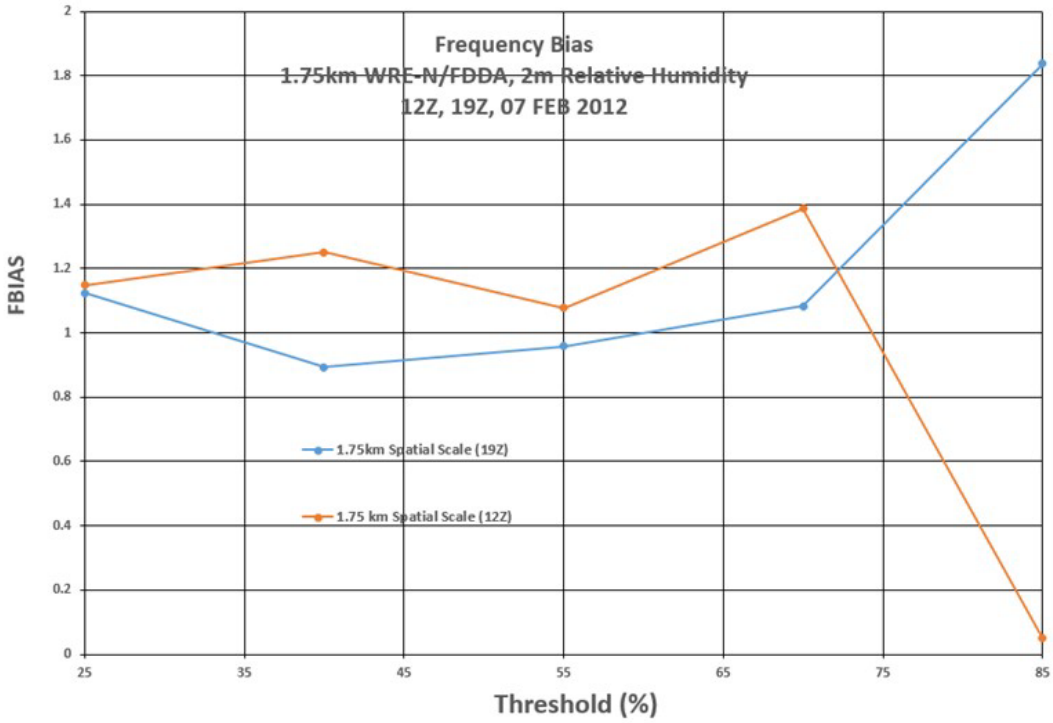

Fig. 22 FBIAS vs. threshold for 1.75-km spatial scale for 2-m-AGL RH at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of FBIAS for RH with threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The superior performance of the WRE–N over the GFS is more obvious with the WRE–N showing less overall bias over a larger range of RH.

Figure 23 shows FBIAS versus threshold value for the nominal range of spatial scales for GFS WIND at preforecast valid time 1200 UTC (lead time = 0 h) for Case 1.

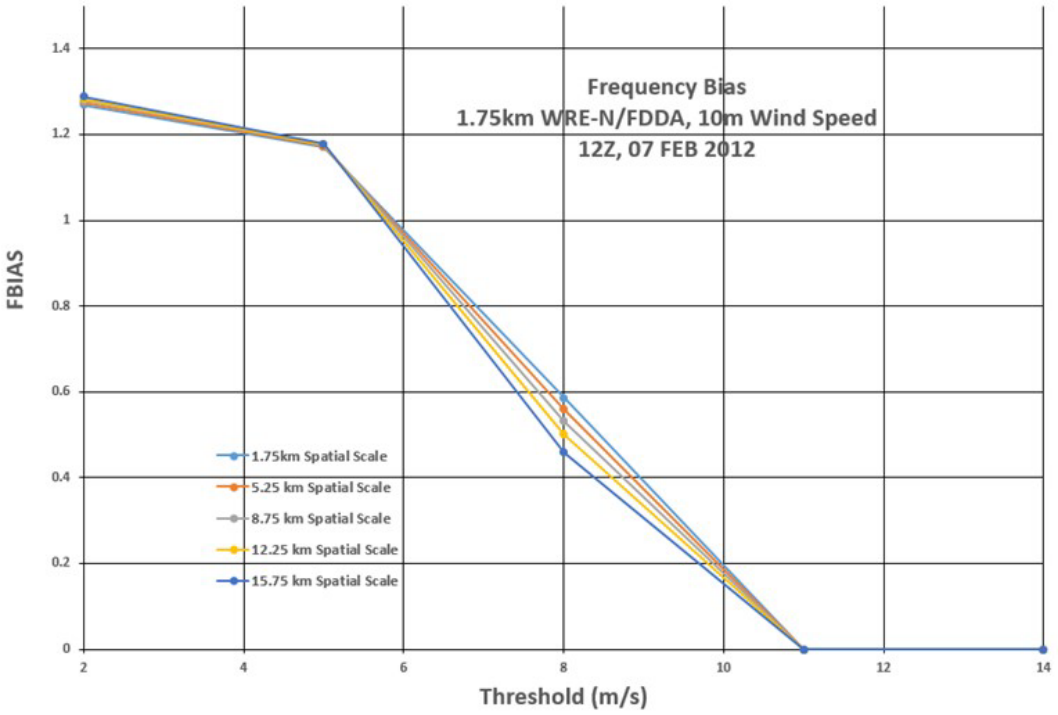

Fig. 23 FBIAS vs. threshold for a range of spatial scales for GFS 10-m-AGL WIND at 1200 UTC for Case 1

The FBIAS for WIND for the GFS shows an over-forecasting trend for the lowest 2 thresholds (2 and 5 m/s), then switches to an under-forecasting trend that increases in magnitude with increasing threshold value. This trend of bias is as expected since the amount of bias at lower thresholds is less than that at higher thresholds. From Fig. 13, this reversal in bias seems to coincide with the changes in object size, which for the lowest 2 thresholds is significantly larger than the objects at higher thresholds. At 2 and 5 m/s, the object areas are roughly 3/4 of the domain and 1/3 of the domain respectively, but at 8 m/s the object size drops to less than 10% of the area. At higher thresholds the bias is shown as 0 because the occurrences of GFS WIND GE 11 and 14 m/s are nil. From Fig. 13, there is no WIND object at these threshold values, which also supports an FBIAS value of 0. There is little evidence of the expected trend—increases in the spatial scale result in a decrease in the FBIAS—as noted by the curves coinciding for all threshold values.

Figure 24 presents FBIAS versus threshold value for the nominal range of spatial scales for WRE–N WIND at forecast valid time 1900 UTC (lead time = 7 h) for Case 1.

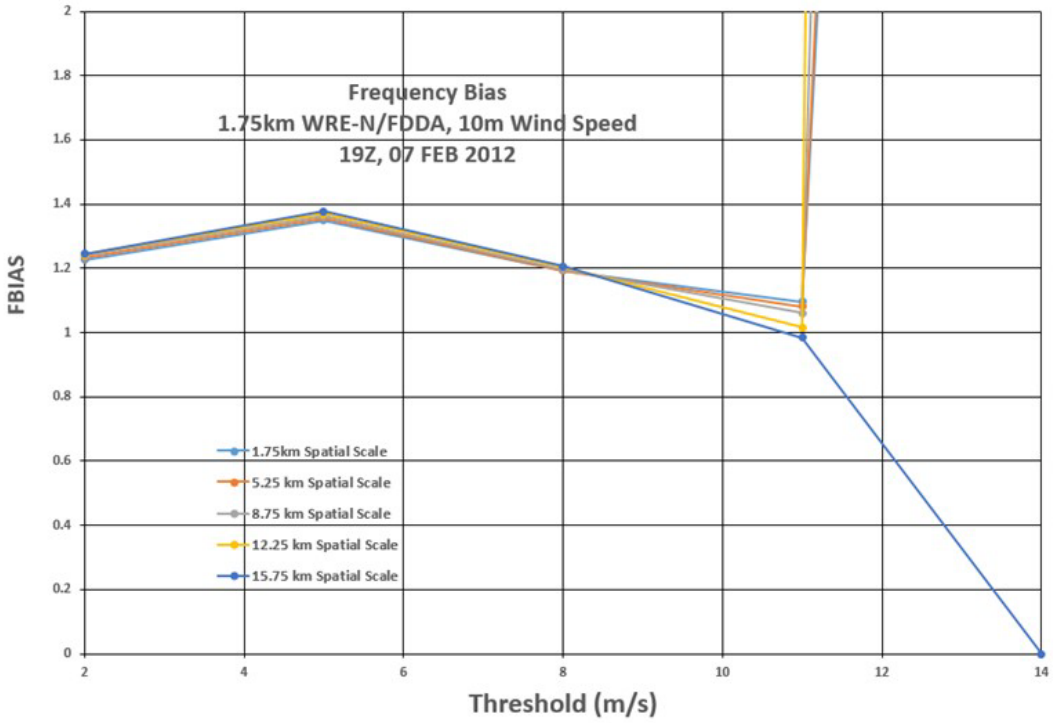

Fig. 24 FBIAS vs. threshold for a range of spatial scales for WRE–N 10-m-AGL WIND at 1900 UTC for Case 1

The FBIAS for WIND for the WRE–N shows an over-forecasting tendency for the lowest 3 thresholds, then reduces to no bias at 11 m/s. The expected trend of increasing bias with increasing threshold value is not strictly present between 2 and 8 m/s. At the 14-m/s threshold, there is a curious separation of the values for bias for the 15.75-km scale from the bias values for the other scales. A check of the data revealed the data value for FBIAS at 15.75 km is "NA" while the FBIAS values for the other scales range from approximately 15 at 1.75 km to approximately 70 at 12.25 km, which are rather extreme values suggesting a large over-forecasting bias. From Fig. 15, for the threshold of 14 m/s there are small objects covering a very small portion of the domain. To confirm this apparent scale-dependent behavior, one would need to compare the occurrences of *observed* WIND GE 14 m/s with that from the shown in Fig. 15, which shows only WRE–N forecast occurrences, in order to explain the difference in the FBIAS calculation. The "NA" value may have resulted from a situation where, considering the larger neighborhood size, the observed occurrences of GE 14 m/s were 0, leading to an undefined FBIAS calculation. On the other hand, at smaller scales the *observed* occurrences of GE 14 m/s was non-zero but very small, leading to a large FBIAS value given the relatively larger number of forecast occurrences of WIND GE 14 m/s. Confirmation of the preceding would require analysis of contingency-table statistics such as the base rate and forecast rate, which provide information on the occurrences in both

fields. There is little evidence of the expected trend—increases in the spatial scale result in a decrease in the FBIAS—as noted by the curves coinciding for most threshold values

Fig. 25 shows FBIAS versus threshold value for 1.75-km spatial scale for model WIND at valid times 1200 and 1900 UTC (lead times = 0 and 7 h) for Case 1.

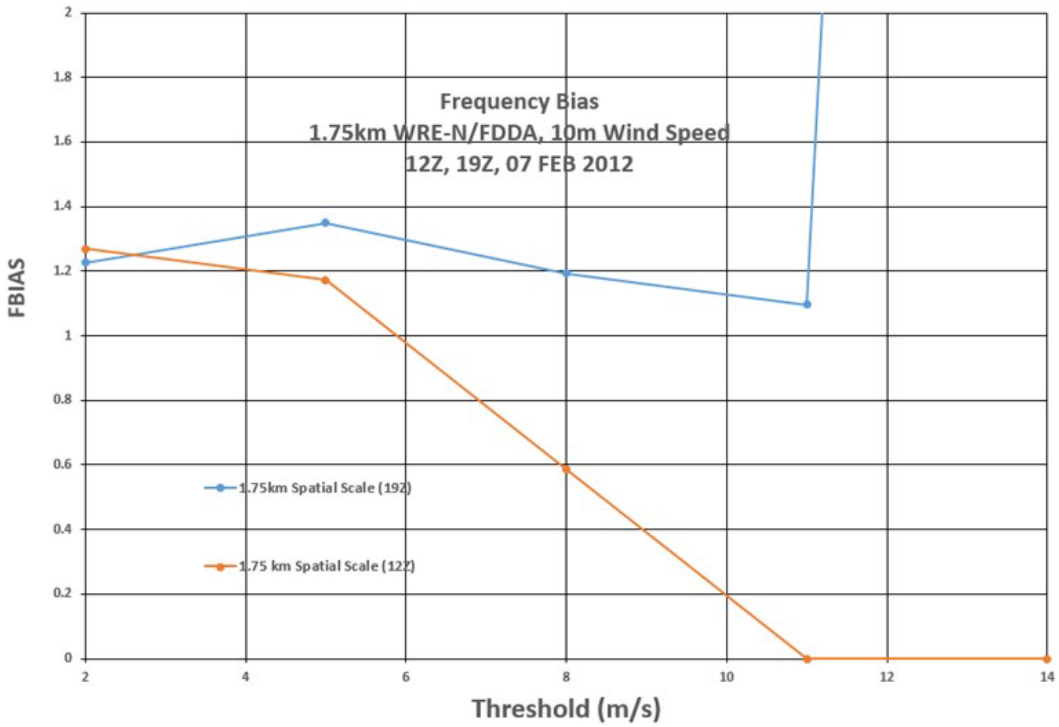

Fig. 25 FBIAS vs. threshold for 1.75-km spatial scale for 10-m-AGL WIND at valid times 1200 (GFS) and 1900 UTC (WRE–N), lead times = 0 and 7 h, for Case 1

For a comparison of the trend of FBIAS for WIND with threshold for 1200 and 1900 UTC, the curves at 1.75-km spatial domain were plotted together. The comparison at the other spatial scales is judged to be similar. The superior performance of the WRE–N over the GFS is more obvious with the WRE–N showing less overall bias over a larger range of WIND. This is especially the case for the threshold values of 8 and 11 m/s. At 14 m/s, the very low number of *observed* occurrences at 1900 UTC and the absence of forecast occurrences at 1200 UTC plays a large role in the difference in the FBIAS values between the GFS and the WRE–N.

5. Conclusions and Final Comments

The minimum coverage method of fuzzy verification, supported by output from the MET Grid-Stat tool, can be applied to the assessment of high-resolution WRE–N model forecasts (Ebert 2008). Furthermore, the tool employs categorical verification techniques that involve the application of threshold values to the spatial fields of continuous meteorological variables, offering the added benefit of a unique type of verification of the WRE–N's ability to simulate objects defined by these thresholds—analogous to objects depicting areas of marginal and unfavorable weather impacts on Army missions and systems, which are the product of the MyWIDA TDA.

It was found that combining the traditional method for verification of categorical forecasts with a nontraditional fuzzy-verification technique offers a more comprehensive approach to assess the ability of the model to predict objects defined by the application of a threshold to a spatial forecast of a continuous variable. This study demonstrated the applicability of the MET Grid-Stat tool for generating aggregated, domain-level categorical contingency-table statistics and scores for continuous meteorological variable fields; also, that CSI and FBIAS statistics will provide a limited, but unique, assessment of model accuracy. However, analysis of additional scores and more cases is necessary in order to clarify relationships between the scores and threshold values as they vary with spatial scale.

This study determined that the choice of the value of the threshold can influence the results. If the threshold is at the high end of the full range of the variable, there can be few or no events occurring, which places limits on calculation of scores and statistics. A reviewer of this study's preliminary results suggested the use of ranges of the variables instead of exceedance criteria in order to extract more information (Reen 2016a). MET Grid-Stat provides additional output of Multi Category Contingency Table statistics that give statistics for bins defined by the threshold values, which may address this concern (NCAR 2013). In addition, the same reviewer suggested that having 2-D plots, such as those generated for this study, is a way to understand the relative coverage of objects defined by the different thresholds; yet, having a quantitative way to determine the percentage of neighborhoods that met the threshold criteria at each scale would add even more information (Reen 2016b).

This study's results suggest that, for the range of spatial scales and the number of thresholds used, the expected trend of decreasing model skill with increasing threshold value was confirmed; however, there is no significant value added by using any particular spatial scale, which runs counter to the expectation that as spatial scale increases, the skill score should increase. A reviewer suggested this

may have been due to the range of neighborhood sizes studied being small compared to the displacement between modeled objects and observed objects and that, perhaps, expanding the neighborhood sizes may reveal such a dependence (Cai 2016).

Analysis of the recurring, decreasing trend observed for CSI and the increasing trend for FBIAS as the threshold value increases, as well as the 2-D plots of the objects defined by the thresholds, suggests an apparent relationship between the size of the objects and the scores. Judging from the scores, particularly at higher threshold values, there is an apparent limitation in the model's ability to simulate objects in continuous variable fields, which seems to be related to the threshold values themselves but may also be related to the area, coverage, and numbers of objects defined by these thresholds. This apparent lack of skill in replicating objects at high threshold values points out a potential weakness of the models that directly impacts the quality of the input to MyWIDA. One reviewer suggested it is probably the reduced size of the objects and the associated increase in displacement errors that makes scoring difficult (Cai 2016). More studies are needed to adequately assess the ability of the models to simulate objects defined using thresholds. Raby and Cai (2016) learned a great deal about the model's ability to simulate objects rendered from continuous fields of meteorological variables by the application of thresholds that can be obtained through the use of the MET MODE tool, which employs nontraditional, object-based methods. Additionally, considerable information about the spatial variability of categorical-verification scores and statistics can be obtained through the use of the MET Series-Analysis tool, which generates spatially distributed, categorical contingency-table statistics and scores for continuous meteorological variable fields (NCAR 2013). Raby and Cai (2016) in their study, which used the MET Series-Analysis tool, found the accuracy of the model, judged from the scores, varies considerably over the domain due to a combination of terrain characteristics and mesoscale variations in the air-mass characteristics. Spatial analysis of more scores and contingency-table statistics is needed to better relate them to terrain and air-mass characteristics. The implications of this variability suggests weather impacts on Army systems and missions vary considerably in space.

A more comprehensive approach combining results from traditional methods with those generated from the application of more nontraditional methods—including object-based methods—may be best for an assessment of the model's skill in predicting fields of a continuous variable that have been filtered by a threshold. Furthermore, the impact of information that such assessments provide about the model's ability to forecast objects on the output of the MyWIDA TDA is of utmost

importance to the Army. This more rigorous approach will certainly require a large amount of data so that statistically significant results can be obtained.

The selection of thresholds with which to generate categorical-verification scores and statistics from the application of both traditional and nontraditional methods will directly impact the extent of useful scores and statistics over the domain. Thus, it is important to include actual system and mission thresholds to more accurately assess the model's ability to predict objects that are meaningful to the Army. That said, actual thresholds' use will result in the significant reduction of numbers of locations and time periods where the atmospheric conditions can provide the range of variable values that encompass these thresholds. The impact of these 2 situations, which are at odds with each other, has to be judged with the understanding that meaningful conclusions about model performance can only come from the analysis of large numbers of cases. So, there is a tradeoff between analyzing data sets for fewer cases in which tactically significant thresholds can be applied and data sets that were developed using thresholds defined by using the actual ranges of the variables present over the domain for the selected case studies. The former presents challenges from lack of statistically significant numbers of cases; the latter presents a challenge of limited application for assessment of the forecast models' ability to forecast objects using mission- and system-specific thresholds.

This preliminary study also documented the first attempt to apply a fuzzy verification method to continuous meteorological variables that have been filtered by a threshold. To the best of the authors' knowledge, this novel approach has never been taken before. Considering that the Army TDAs mostly rely on critical thresholds in continuous variables (e.g., temperature, relative humidity, and wind speed) to issue warnings that might affect Army operations, it is imperative to evaluate the impact of forecast accuracy on these TDAs. By employing both traditional and nontraditional forecast-evaluation methods, such as those demonstrated in this study, a more complete picture of model-forecast performance can be gleaned by analyzing a large amount of forecast data. In that direction, future work will focus on more statistics and, most importantly, more cases so that statistically significant results can be obtained.

Finally, a Geographic Information System (not extensively used in atmospheric sciences) should be exploited for its ability to contextualize and analyze geospatial information—terrain type/slope, land-use effects, and other spatial and temporal variables—as explanatory metrics in model assessments (Smith et al. 2016a; Smith et al. 2016b; Smith et al. 2015). This technique has demonstrated considerable promise as an important new tool that, in addition to the methods described in this study, offers a comprehensive approach to model verification.

6. References

Brandt J, Dawson L, Johnson J, Kirby S, Marlin D, Sauter D, Shirkey R, Swanson J, Szymber R, Zeng S. Second generation weather impacts decision aid applications and web services overview. White Sands Missile Range (NM): Army Research Laboratory (US); 2013 July. Report No.: ARL-TR-6525.

Cai H. Personal communication. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Mar 15.

Cai H, Dumais RE. Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis. Wea and Forecasting. 2015;30:1451–1468.

Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocernich M, Damrath U, Ebert EE, Brown BG, Mason S. Forecast verification: current status and future directions. Meteo App. 2008;15(1):3–18.

Chen F, Dudhia J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part II: preliminary model validation. Mon Wea Rev. 2001a;129:587–604.

Chen F, Dudhia J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon Wea Rev. 2001b;129:569–585.

Daniels TS, Moninger WR, Mamrosh RD. Tropospheric airborne meteorological data reporting (TAMDAR) overview. Preprints, 10th Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface; 2016 Sep 1; Atlanta (GA): American Meteorological Society [accessed 2016 Aug 2]. http://ams.confex.com/ams/pdfpapers/104773.pdf.

De Pondeca MSFV, Manikin GS, DiMego G, Benjamin SG, Parrish DF, Purser RJ, Wu W-S, Horel JD, Myrick DT, Lin Y, et al. The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: current status and development. Wea Forec. 2011;26:593–612.

Deng A, Stauffer D, Gaudet B, Dudhia J, Hacker J, Bruyere C, Wu W, Vandenberghe F, Liu Y, Bourgeois A. Update on the WRF-ARW end-to-end multi-scale FDDA system. Paper presented at: 10th WRF Users' Workshop, National Center for Atmospheric Research, 2009 Jun 23–26; Boulder, CO.

[DTC] Developmental Testbed Center. MET online tutorial for METv3.0: COPYGB functionality. Boulder (CO): National Oceanic and Atmospheric Administration; [accessed 2016 Jul 27]. http://www.dtcenter.org/met/users/support/online tutorial/METv3.0/copygb/index.php.

Dudhia J. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. J Atmos Sci. 1989;46:3077–3107.

Dumais R. Personal communication. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Mar 22.

Dumais R, Kirby S, Flanigan R. Implementation of the WRF four-dimensional data assimilation method of observaion nudging for use as an ARL weather running estimate-nowcast. White Sands Missile Range (NM): Army Research Laboratory (US); 2013 Jun. Report No.: ARL-TR-6485.

Dumais RE, Reen BP. Data assimilation techniques for rapidly relocatable weather research and forecasting modeling. White Sands Missile Range (NM): Army Research Laboratory (US); 2013 Jun. Report No.: ARL-TN-0546.

Dumais RE, Raby JW, Wang Y, Raby YR, Knapp, D. Performance assessment of the three-dimensional wind field Weather Running Estimate-Nowcast and the three-dimensional wind field Air Force Weather Agency weather research and forecasting wind forecasts. White Sands Missile Range (NM): Army Research Laboratory (US); 2012 Dec. Report No.: ARL-TN-0514.

Dumais Jr. RE, Henmi T, Passner J, Jameson T, Haines P, Knapp D. A mesoscale modeling system developed for the U.S. Army. White Sands Missile Range (NM): Army Research Laboratory (US); 2004. Report No.: ARL-TR-3183.

Ebert E, Wilson L, Weigel A, Mittermaier M, Nurmi P, Gill P, Göber M, Joslyn S, Brown B, Fowler T, et al. Progress and challenges in forecast verification. Meteo App. 2013;20(2):130–139.

Ebert E. Fuzzy verification of high resolution gridded forecasts: a review and proposed framework. Meteo App. 2008;15:51–64.

[EMC] Environmental Modeling Center. The GFS Atmospheric Model. Washington (DC): National Weather Service-National Centers for Environmental Prediction; 2003 Nov. NCEP Office Note No.: 442.

Google Earth. Mountain View, CA; 2016 [accessed 2016 Aug 24]. http://maps.google.com/help/terms_maps.html

Hong SY, Dudhia J, Chen SH. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Wea Rev. 2004;132:103–120.

Janjić ZI. The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon Wea Rev. 1994;122:927–945.

Jolliffe IT, Stephenson DB. Forecast verification: a practitioner's guide in atmospheric science. 2nd ed. Hoboken (NJ): John Wiley and Sons; 2012.

Kain JS. The Kain-Fritsch convective parameterization: an update. J App Meteo. 2004;43:170–181.

Liu Y, Bourgeois A, Warner T, Swerdlin S, Hacker J. Implementation of observation-nudging based FDDA into WRF for supporting ATEC test operations. Paper presented at: 6th WRF/15th MM5 Users' Workshop, National Center for Atmospheric Research; 2005 Jun 27–30; Boulder, CO.

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geoph Res Atmo. 1997;102:16663–16682.

[NCAR] National Center for Atmospheric Research. Model evaluation tools version 4.1 (METv4.1), user's guide 4.1. Boulder, CO; 2013 May.

[NOAA] Meteorological assimilation data ingest system (MADIS). College Park (MD): National Oceanic and Atmospheric Administration [accessed 2016 Jul 27]. http://madis.noaa.gov.

[NRC] National Research Council. When weather matters: science and service to meet critical societal needs. Washington (DC): The National Academies Press; 2010.

Raby JW, Cai H. Verification of spatial forecasts of continuous meteorological variables using categorical and object-based methods. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Aug. Report No.: ARL-TR-7751.

Reen B. Personal communication. 2016a. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Mar 22.

Reen B. Personal communication. 2016b. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Mar 16.

Reen BP, Schmehl KJ, Young GS, Lee JA, Haupt SE, Stauffer DR. Uncertainty in contaminant concentration fields resulting from atmospheric boundary layer depth uncertainty. J App Meteo Clim. 2014;53:2610–2626.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang X-Y, Wang W, Powers JG. A description of the advanced research WRF version 3. Boulder (CO): National Center for Atmospheric Research (US); 2008 Jun. NCAR Technical Note No.: TN–475+STR.

Smith J. Personal communication. White Sands Missile Range (NM): Army Research Laboratory (US); 2016 Mar 14.

Smith JA, Foley TA, Raby JW, Reen B. Investigating surface bias errors in the Weather Research and Forecasting (WRF) model using a Geographic Information System (GIS). White Sands Missile Range (NM): Army Research Laboratory (US); 2015 Feb. Report No.: ARL-TR-7212.

Smith JA, Foley TA, Raby JW, Reen BP, Penc RS. 2016a. Case study applying GIS tools to verifying forecasts over a domain. Paper presented at: 96th annual American Meteorological Society Meeting, 23rd Conference on Probability and Statistics in the Atmospheric Sciences; 2016 Jan 10–14; New Orleans, LA. Paper No.: 13.3.

Smith JA, Raby JW, Foley TA, Reen BP, Penc RS. 2016b. Case study applying GIS tools to verifying forecasts over a mountainous domain. Paper presented at: 17th Mountain Meteorology Conference, American Meteorological Society. 2016 Jun 27–July 1; Burlington, VT. Paper No.: 2.5.

Stauffer DR, Seaman NL. Multiscale four-dimensional data assimilation. J App Meteoy. 1994;33:416–434.

Vaucher G, Raby J. Assessing high-resolution weather research and forecasting (WRF) forecasts using an object-based diagnostic evaluation. White Sands Missile Range (NM): Army Research Laboratory (US); 2014 Feb. Report No.: ARL-TR-6843.

Wilks DS. Statistical methods in the atmospheric sciences. 3rd ed. Oxford (UK): Academic Press; 2011.

List of Symbols, Abbreviations, and Acronyms

ACARS	Aircraft Communications, Addressing, and Reporting System
ACC	accuracy score
AGL	above ground level
ARL	US Army Research Laboratory
ARW	Advanced Research Weather Research and Forecasting model
BASER	base rate
CSI	critical success index
DPT	dew-point temperature
FAR	false-alarm ratio
FBIAS	frequency bias
FDDA	Four-Dimensional Data Assimilation
FMEAN	mean forecast value
GE	greater than or equal to
GFS	Global Forecast System
GRIB	Gridded Binary format Edition 1
GRIB2	Gridded Binary format Edition 2
GSD	Global Systems Division
GSS	Gilbert Skill Score
hPa	hectopascal
K	degrees Kelvin
LAPS	Local Analysis and Prediction System
MADIS	Meteorological Assimilation Data Ingest System
MCTC	Multicategory Contingency Table Counts
MET	Model Evaluation Tools
MODE	Method for Object-Based Diagnostic Evaluation
MYJ	Mellor–Yamada–Janjic

MyWIDA	My Weather Impacts Decision Aid
NCAR	National Center for Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PBL	planetary boundary layer
PODY	probability of detection—hit rate
RH	relative humidity
RRTM	Rapid Radiative Transfer Model
RTMA	Real-Time Mesoscale Analysis
TAMDAR	Tropospheric Airborne Meteorological Data Reporting
TDA	tactical decision aid
TKE	turbulent kinetic energy
TMP	temperature
UTC	coordinated universal time
WIND	wind speed
WRE–N	Weather Running Estimate–Nowcast
WRF	Weather Research and Forecasting
WRF–ARW	Weather Research and Forecasting, Advanced Research WRF
2-D	2-dimensional

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO L IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
12 (PDFs)	US ARMY RSRCH LAB ATTN RDRL CIE M J RABY H CAI G VAUCHER D KNAPP J SMITH J PASSNER R PENC S KIRBY R DUMAIS T JAMESON B REEN B MACCALL
1 (PDF)	US ARMY RSRCH LAB ATTN RDRL CIE P CLARK
3 (PDFs)	US ARMY RSRCH LAB ATTN RDRL CIE D R RANDALL S O'BRIEN J JOHNSON
1 (PDF)	US NAVY RSRCH LAB DR J MCLAY
1 (PDF)	US AIR FORCE 557TH WEATHER WING R CRAIG
1 (PDF)	DCGS-A WEATHER EET LEAD J CARROLL
3 (PDF)	UCAR T FOWLER J H GOTWAY B BROWN
1 (PDF)	USAICOE J STALEY